

AGGACAGAGC TCACCTTCTT GATCAAGAGC GGCTGGATGC TCTACAGCGC ATGGGCGATC 480  
 GAGTTCTCTG AGGCGCGAGC CAACCTGA 507

## (2) INFORMATION FOR SEQ ID NO:95:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 168 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:95:

Met Lys Met Val Lys Ser Ile Ala Ala Gly Leu Thr Ala Ala Ala Ala 1  
 5 10 15  
 Ile Gly Ala Ala Ala Ala Gly Val Thr Ser Ile Met Ala Gly Gly Pro 20  
 25 30  
 Val Val Tyr Gln Met Gln Pro Val Val Phe Gly Ala Pro Leu Pro Leu 35  
 40 45  
 Asp Pro Ala Ser Ala Pro Asp Val Pro Thr Ala Ala Gln Leu Thr Ser 50  
 55 60  
 Leu Leu Asp Ser Leu Ala Asp Pro Asn Val Ser Phe Ala Asn Lys Gly 65  
 70 75 80  
 Ser Leu Val Gln Gly Gly Ile Gly Gly Thr Gln Ala Arg Ile Ala Asp 85  
 90 95  
 His Lys Leu Lys Lys Ala Ala Gln His Gly Asp Leu Pro Leu Ser Phe 100  
 105 110  
 Ser Val Thr Asn Ile Gln Pro Ala Ala Gly Ser Ala Thr Ala Asp 115  
 120 125  
 Val Ser Val Ser Gly Pro Lys Leu Ser Ser Pro Val Thr Gln Asn Val 130  
 135 140  
 Thr Phe Val Asn Gln Gly Gly Trp Met Leu Ser Arg Ala Ser Ala Met 145  
 150 155 160  
 Gln Leu Leu Gln Ala Ala Gly Asn 165

## (2) INFORMATION FOR SEQ ID NO:96:

- (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 500 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:96:

CTGTGCAATG TCGTTCACCG TCGGGGCGCG GTTCGGCTCC GCAGATCCCG TGGACCGGCT	60
CATTAAACCC ACGTCGAATT ACGGTCGAGT AGTAGCTGCG CTCAAGCGCG CCGATCCGGG	120
GGCTACCGCA CATTGCAAGC CCGCACCGGT GCGGCAGTCC TATTGCGGCA ATTTGCTGCG	180
CGCACCGGCA CCTGAGCGCG CTGCAATGGC CGGCGATTGC CAAGCTGTTC CGGGGGCGGC	240
ACAGTACATC GCGCTCTGCG AGTCGTTGCG CCGCTCTGCG AACAGTATT AACCGCATGC	300
GGCGGCGATC CCGGCGACCG AGATCTGCGC CCGGCGTACG CCAGATTGCG CCGCTCTGCA	360
ACCGGCGGCA TCGCGGACCG CCGGCTGCTC GCGGCGGCTA GCGGAGATTG CCGGCTCTCT	420
CGACGCGGCG CATCTGTGCG CGAATTCTTC CAGCGCGGCG GATCGCATAG TTGTAGAGCG	480
GGCGGCGGCG CCGTTCGAGCT	500

(zi) INFORMATION FOR SEQ ID NO:97:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 98 amino acids  
 (B) TYPE: amino acids  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:97:

Val Ala Met Ser Leu Thr Val Gly Ala Gly Val Ala Ser Ala Asp Pro	
1 5 10 15	
Val Asp Ala Val Ile Asn Thr Thr Cys Asn Tyr Gly Gln Val Val Ala	
20 25 30	
Ala Leu Asn Ala Thr Asp Pro Gly Ala Ala Ala Gln Phe Asn Ala Ser	
35 40 45	
Pro Val Ala Gln Ser Tyr Leu Arg Asn Phe Leu Ala Ala Pro Pro Pro	
50 55 60	

Gln Arg Ala Ala Met Ala Ala Gln Leu Gln Ala Val Pro Gly Ala Ala  
65 70 75 80

Gln Tyr Ile Gly Leu Val Gln Ser Val Ala Gly Ser Cys Asn Asn Tyr  
85 90 95

## (2) INFORMATION FOR SEQ ID NO:98:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 154 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:98:

ATGACAGAGC AGCACTGGAA TTTCGCGGT ATGAGAGGCC CAGCAGCCG AATCCAGGGA 60  
 AATGTCACCT CATTCAATTC CTTCCTTGA: CAGGAGAGC AGTCCCTGAC CAAGCTCGCA 120  
 GCGGCTCTGG GCGTATGGG TTGGGAGGC TACC 154

## (2) INFORMATION FOR SEQ ID NO:99:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 51 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:99:

Met Thr Glu Gln Gln Trp Asn Phe Ala Gly Ile Gln Ala Ala Ala Ser  
1 5 10 15  
 Ala Ile Gln Gly Asn Val Thr Ser Ile His Ser Leu Leu Asp Glu Gly  
20 25 30  
 Lys Gln Ser Leu Thr Lys Leu Ala Ala Ala Trp Gly Gly Ser Gly Ser  
35 40 45  
 Glu Ala Tyr  
50

## (2) INFORMATION FOR SEQ ID NO:100:

- (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 260 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

CGCTTCGGCA CTTCAGGTA ACTATGAAAG TCGCTTCGG NCTGGAGGAT TCTGAACCT	60
TCAAGGCGCG CGGATACACT AGGTGCTCA TTAAGCGACT TTTCAGAC ATCTGACGC	120
CTCGAAGAG CGGCACAGC GAGGTGGCT CGGCGAGGC GCTGCTGCT AAATCCTGA	180
GACATTCG CGGAGGCGG TACAGGAAG TCGGTGCTA ATTCGCGCG TATCTGCTG	240
ACCTGTTCG TCTGAGCGG GACGAGCGG TCTCTGAGCT CG	262

(12) INFORMATION FOR SEQ ID NO:101:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3058 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

GATCGTCCC GTGAGATGC TCGGCGCTT TGAGATGGA GTGACCTGT CTTCCTGAT	60
GGATACCGA GAGATGTGG CGGCGCGGC TGACACCTG CAGGACATG GTGCTAGGAC	120
TGTGACTAG AATCTCGTG CGGCGCGGC GACGACTGG GTGATGCGC CCGCTGCGA	180
TGAGGTGTC GCGGTGACT CGGCGCGCT CGGCGCAT GCGGCGATG ATGACTTGT	240
GAGCGCTCG GCTGCTGCA TTGATGACA GTTCTGCGC ACCTTGCAG GAGGCGGAG	300
CTGCTATGC GCACTGAG TCGCATCG CGCGCGCGC AGCTAAGCA GAGGATCG	360
CGCGGAAA CCAGAGAA TGGAGACCG TATGATGGA TTTCGCGGC TTACACCGG	420
AGATCAACT CCGAGGATG TGGGCGGCC CGGCTTGGC CTCCTGCTG CGCGCGCTC	480
AGATGTGCA CAGCTGCGC ACTGACTGT TTTCGCGGC GTGCGCTTT CAGTCTGCG	540
TCTGGGTTT GAGCTGCGG TCGTGTGTA GTTCGTGCG GGTCTGATG GTGCGCGCG	600

CGTCGGCTTA	TGTGGGCTG	ATGAGGCTCA	CCGGGGGGA	GGCGAGCTG	ACGGCGGCG	660
AGGTCCGGT	TGCTCGGCG	GCTACGAGA	CGGCTATCG	GTGACGGT	CGCGCGCGG	720
TGATGGCGA	GAACGCTCG	GAAGTATGA	TTCTGATAG	GACCAAGCT	TTGGGCGAA	780
ACACCTCGG	GATGGCGTC	TACGAGCGG	AATACGGGA	GATTCGCGC	CAAGACGGG	840
CGCGATGTT	TGGCTAGCG	GCGCGAGCG	CGACGGGAC	GGGACGTTG	CTGGCTTTG	900
AGGAGGGCG	GGAGATGAC	AGCGGCGTG	GGCTTTTGA	GCAAGCGCG	CGGCTGAGG	960
AGGCTCGGA	CACGCGCGG	GGAACGAGT	TGATGAGCA	TGTGGCGAG	GGCTTGCAG	1020
AGCTGGCGA	GCGCAGCGG	GCGACGAGC	CTTCTTCAA	GCTGGTGGG	CTGTGAGCA	1080
CGGTCTCGC	GATCGGCTG	CGATCGGCA	ACATGGTGT	GATGGCGAG	AGCGATGAT	1140
CGATGACAA	CTGAGCTCG	TGATGAGCA	AGCGCTGAG	CTGATGTTT	AGGGGCTTT	1200
CTCGCGCGC	GGCGCGCGG	GCTTTCGAA	CGCGCGGCA	AAACGGGCT	CGGGCGATG	1260
GTGCACTGG	CAGTCTCGT	GCTTCTCGG	GTGCGCGCG	TGGGTGGCG	GCGAAGTGG	1320
GTGCGCGCG	CTCGGCTCG	TGTTCTCGG	TGCGVAGG	CTGCGCGCG	GCGACGAGG	1380
CAGTCACTG	GGCGCGCGG	GCGTCTCGG	TGCGCGGCT	GCGCGCGCG	GCGGAGAGG	1440
GGCGCGCGG	GATCTGGCG	GCGCTGCGG	TGCGCGGCT	GCGCGCGCG	GCGGTGCTG	1500
GCTCACTGG	TGTCTGCTG	GTGCGCGCG	GAGCTATGT	GATCGCGAT	TGCGCGCGG	1560
CGGCTAGGA	GAGCGCGCG	AGCTTCTGT	TATTTGAGC	GTGCTGCGG	CTGCTGCTG	1620
TTGCTAGCG	GCTTATGCA	AGCTTATG	TGCTGAGCA	GTGCTAGCA	TTAGGTGCG	1680
GTGCACTAG	CAGACGCGG	AGCTGCGCT	AGCTTTATG	AGCTATGCG	AGCTATGCG	1740
CGACATGCG	GGCGCTTTG	AGCTGCGCT	CGACAGCTG	GAGGAGAGG	CTGCTGCGAT	1800
GTGCGCTCG	GCGGAGAGG	TTTCTGCTG	GCGCTGAGT	GCGCTGCGG	AGCGGAGCT	1860
GCTAGACAG	ATGCGCGAG	TGATGAGG	GTGCTGAG	ATGCTGAG	TGCTGAGG	1920
GCTGCTGAG	GCTGCTG	GCGAGCTG	AGCTTATG	AGCTTATG	AGCTTATG	1980
GCGCTGCTG	AGCTTATG	GCTGCTG	AGCTTATG	AGCTTATG	AGCTTATG	2040
AGCTTATG	TGAGCTGAG	GTGCTGAG	GCGCTGAG	AGCTTATG	GCGCTGAG	2100
GCGCTGAG	GCGCTTATG	GCGCTGAG	GTGCTGAG	TGATGCTG	TGCTTATG	2160
GCGCTGAG	TTTCTGAG	GCGCTTATG	GCGCTTATG	AGCTTATG	TGCTTATG	2220
GCGCTTATG	TGCTTATG	GTGCTGAG	GCGCTTATG	AGCTTATG	GCGCTTATG	2280

```

GCCGCGAACA ACATGGCGCA AAGGACAGC GCGGTGGGT CAGGTGGGC CTACACCGC 2340
GCCAAGCGCA GGAAGTGGT GACAGAGTA AGTTCTCGC GTGATCTTC GAGTGGCAGT 2400
CTAATGGTC AGTGTCTGG TGTAGTGGT TTGCTGCTG GGGGGTCTT CGGTGCTGGT 2460
CAGTGTGCT CCGCTCGAG TGAGGACCTC GAGGCCAGG TAGCGCGTC CTTCATCCA 2520
TTGCTCTGT TTTTGGCGA GAGGGGTTC CACGAGGAG ATGATCGAG GAGGCTCGG 2580
GAAGATGCT ACCAGCTGG TTGGCGGTC TACCTCTCG TTGAGCGCT CCTGGGGCTT 2640
GTTCAGCAG ATTTGGGCT AGATCTGCT GGGGAGGCC GTGAGCGCA GAGTGGGT 2700
GGGGCGGTA TAGAGTGGT GGGGCTGCG GGGGAGTTG TGGTCAAGG GGTGAGTAC 2760
GGGATATAT TGGGAACAA GTGATTCAG GTGGGCTCG TGTAGTGG AGTGAGCAG 2820
GCTGGGAGC CACCGCAGG AGGCTTCGG GTTGGCTGC ATCAGATTG CTGCTATTC 2880
GCTTCTCGA CCGTGTGAG CCGTGGCGG CAGGTTGAG TCGATCGAG CACGAGGCT 2940
GGGTGGGGT TCGGTGTGA CAGCGCGAC CCGGACAGG GCGGGGCGA CCGGTGGCG 3000
GAAGAGTGC AGCGAGCGG CCGCTCTCT GCGGAGGTC AGTGGATGC CCGGATC 3058

```

## (2) INFORMATION FOR SEQ ID NO:102:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 391 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

```

Met Val Asp Phe Gly Ala Leu Pro Pro Glu Ile Asn Ser Ala Arg Met
1           5           10          15
Tyr Ala Gly Pro Gly Ser Ala Ser Leu Val Ala Ala Ala Gln Met Trp
20          25          30
Asp Ser Val Ala Ser Asp Leu Phe Ser Ala Ala Ser Ala Phe Gln Ser
35          40          45
Val Val Trp Gly Leu Thr Val Gly Ser Trp Ile Gly Ser Ser Ala Gly
50          55          60
Leu Met Val Ala Ala Ala Ser Pro Tyr Val Ala Trp Met Ser Val Thr
65          70          75          80

```

Ala Gly Gln Ala Glu Leu Thr Ala Ala Gln Val Arg Val Ala Ala Ala  
 85 90 95  
 Ala Tyr Glu Thr Ala Tyr Gly Leu Thr Val Pro Pro Pro Val Ile Ala  
 100 105 110  
 Glu Asn Arg Ala Glu Leu Met Ile Leu Ile Ala Thr Asn Leu Leu Gly  
 115 120 125  
 Glu Asn Thr Pro Ala Ile Ala Val Asn Gln Ala Glu Tyr Gly Glu Met  
 130 135 140  
 Trp Ala Gln Asp Ala Ala Ala Met Phe Gly Tyr Ala Ala Ala Thr Ala  
 145 150 155 160  
 Thr Ala Thr Ala Thr Leu Leu Pro Phe Glu Gln Ala Pro Glu Met Thr  
 165 170 175  
 Ser Ala Gly Gly Leu Leu Glu Gln Ala Ala Val Gln Glu Ala Ser  
 180 185 190  
 Asp Thr Ala Ala Ala Asn Gln Leu Met Asn Asn Val Pro Gln Ala Leu  
 195 200 205  
 Gln Gln Leu Ala Gln Pro Thr Gln Gly Thr Thr Pro Ser Ser Lys Leu  
 210 215 220  
 Gly Gly Leu Trp Lys Thr Val Ser Pro Ala Arg Ser Pro Ile Ser Asn  
 225 230 235 240  
 Met Val Ser Met Ala Asn Asn His Met Ser Met Thr Asn Ser Gly Val  
 245 250 255  
 Ser Met Thr Asn Thr Leu Ser Ser Met Leu Lys Gly Phe Ala Pro Ala  
 260 265 270  
 Ala Ala Ala Gln Ala Val Gln Thr Ala Ala Gln Asn Gly Val Arg Ala  
 275 280 285  
 Met Ser Ser Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu Gly Gly Gly  
 290 295 300  
 Val Ala Ala Asn Leu Gly Arg Ala Ala Ser Val Gly Ser Leu Ser Val  
 305 310 315 320  
 Pro Gln Ala Trp Ala Ala Ala Asn Gln Ala Val Thr Pro Ala Ala Arg  
 325 330 335  
 Ala Leu Pro Leu Thr Ser Leu Thr Ser Ala Ala Gln Arg Gly Pro Gly  
 340 345 350  
 Glu Met Leu Gly Gly Leu Pro Val Gly Gln Met Gly Ala Arg Ala Gly  
 355 360 365

Gly Gly Leu Ser Gly Val Leu Arg Val Pro Pro Arg Pro Tyr Val Met  
 370 375 380

Pro His Ser Pro Ala Ala Gly  
 385 390

## (2) INFORMATION FOR SEQ ID NO:103:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1725 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (x) SEQUENCE DESCRIPTION: SEQ ID NO:103:

GAGCTGCGCC	GGGGGGCTCC	AGGGCTGGAG	CGCTGTGAGT	TTTGATCTCG	GGTCAAGCTG	60
AGGTGCTCTG	GGGTGTGGGG	GGGGTGGATC	CAGACTGAGT	GGGGCTCTTT	AGTGCACCTA	120
ATTTCGTTGA	AGTGGCTCTG	AGTTATAGGA	CTTCAGGATT	GGTTAATGTA	GGTTTCACCC	180
CGTGTGGGG	TGATTTTTGG	CGGAGCAGTC	GTGGCAGAGC	CTTGCGCTCC	GGGTCAAGCG	240
GGGATGACGA	TGGTTTACT	AGCAATCAAT	CTTGGCTCC	GGTGGCAGTG	CTGGGCTTAA	300
ATGAGGARGA	GGAGCGGTGT	CTTTCAGCTC	GGACGGGAG	ATGTTGGGGG	CGGGGGCTGG	360
CGAAGTTGCT	TGGTGGGG	CAAGCTGAA	GGCTAGCAAT	GGGGGGGAG	CGGTGGGAG	420
GAGTGGGGTG	GTGGGGGGGG	CTGGGAGGGA	GGTGTGGCTG	CTGGTTGAGA	CACAATTCTG	480
TAGGATGTGG	GGGAGCTATC	AGAGGGGGGG	GGGCAAGGCG	GGGTGATCC	ATGAGCAGTT	540
TGTGACCGAG	CTGGGACCGA	GGGCTATTTC	ATATGGGAG	ACGGAGGGGG	CGAAGCTGT	600
GTTCAGCTGC	TGGTGTACT	GAGGTATTTC	GAGCGGAAGG	ATTATCGAAG	TGGTGGATTI	660
CGGGGGTTTA	CGGGGGGAGA	TCAACTCGGC	GAGGATGTAT	GGGGGGGGGG	GTTCGGGCTC	720
GGTGTGGTGG	GGGGGAGAGA	TGTGGGAGAG	GTGGGGAGCT	CAGTTGTATT	GGGGGGGGTC	780
GGGTTTTCAG	TGGTGGTGTG	GGGGTCTGAG	GGTGGGGTGG	TGGATAGATT	GGTGGGGGGG	840
TGTGATGGGG	GGGGGGGGGG	GGGGTGTGTT	GGGGGAGAG	AGGCTGAGGG	GGGGGAGGGC	900
CGAGCTGAGG	GGGGGGGGGG	TGGGGTTTCC	TGGGGGGGGC	TAGGAGAGAG	GGTATAGGCT	960
GAGGCTGGCC	GGGGGGGGGG	TGGGGGGGAG	GGTACCGGAG	CTGATGAGGG	TGAGCGGAGC	1020
CACTCTCTCG	GGGGGAGAGA	GGGGGGGAGT	GGGGGGGAGT	GGGGGGGAGT	AGAGCGAGAT	1080



```

GTGGGGCCAA GACGGGGAGC CGATGTATGC CTACGGGACC AGGGGGGCGA GGGGGACCGA 1140
GGGGTTGGTG CGGTTCGAGG AGGGGGGCT GATCGGCAAC GGGGGCGGGC TGGTTGAGCA 1200
GGGGGTGGGG CTGGAGGAGG CGATCGGACG GGGGGGGGGG AAGCGATTGA TGACAAATGT 1260
GGGCCAAGCG CTGCAACAGC TGGGGCAAGC AGGGGAGGGG GTCTGTCCTT GTTGAAAGCT 1320
GGGTGGGGTG TGGAGCGGGG TCTCGGGGCA TGTGTGGGAG CTAGGCAAGC TCAATTGGAT 1380
AGCAACAAC CAGATGTGTA TGATGGGAGC GGGTGTGTGC ATGACCAACA GTTGCACTTC 1440
GATGTTAAAG GCGTATATTC CGGGGGGGCC TCAGGGGCTG GAGGGGGGGG GGGGAAAGCG 1500
GGTCTGGGCG ATGAGGCTGC TGGGAGGCGA GCTGGGTTCG TGGCTGGGTT CTTCGGGTCT 1560
GGGCGCTGCG GTGGGGGCGA ACTGGGTTCG GGGGGGCTCG GTGGGTTCGT TGTGGGTGCG 1620
GGCAGCATGG GGGGGGCGA AGGGGGGGCT GATGGGGGGG GGGGGGGGGG TGGGGCTGAC 1680
GAGGCTGAGC AGGGGGGGC AAGGGGGGGC GGGGGGAGAT GTGGG 1720

```

(2) INFORMATION FOR SEQ ID NO:104:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 309 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: Linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:104:

```

Val Val Asp Phe Gly Ala Leu Pro Pro Gln His Asn Ser Ala Arg Met
 1           5           10           15
Tyr Ala Gly Pro Gly Ser Ala Ser Leu Val Ala Ala Ala Lys Met Trp
          20           25           30
Asp Ser Val Ala Ser Asp Leu Phe Ser Ala Ala Ser Ala Phe Gln Ser
          35           40           45
Val Val Trp Gly Leu Thr Val Gly Ser Trp His Gly Ser Ser Ala Gly
          50           55           60
Leu Met Ala Ala Ala Ala Ser Pro Tyr Val Ala Trp Met Ser Val Thr
          65           70           75           80
Ala Gly Gln Ala Gln Leu Thr Ala Ala Gln Val Arg Val Ala Ala Ala
          85           90           95

```

Ala Tyr Glu Thr Ala Tyr Arg Leu Thr Val Pro Pro Pro Val Ile Ala  
 100 105 110  
 Glu Asn Arg Thr Glu Leu Met Thr Leu Thr Ala Thr Asn Leu Leu Gly  
 115 120 125  
 Glu Asn Thr Pro Ala Ile Glu Ala Asn Glu Ala Ala Tyr Ser Glu Met  
 130 135 140  
 Trp Gly Glu Asp Ala Glu Ala Met Tyr Gly Tyr Ala Ala Thr Ala Ala  
 145 150 155 160  
 Thr Ala Thr Glu Ala Leu Leu Pro Phe Glu Asp Ala Pro Leu Ile Thr  
 165 170 175  
 Asn Pro Gly Gly Leu Leu Glu Glu Ala Val Ala Val Glu Glu Ala Ile  
 180 185 190  
 Asp Thr Ala Ala Ala Asn Glu Leu Met Asn Asn Val Pro Glu Ala Leu  
 195 200 205  
 Glu Glu Leu Ala Glu Pro Ala Glu Gly Val Val Pro Ser Ser Lys Leu  
 210 215 220  
 Gly Gly Leu Trp Thr Ala Val Ser Pro His Leu Ser Pro Leu Ser Asp  
 225 230 235 240  
 Val Ser Ser Ile Ala Asn Asn His Met Ser Ser Ser Gly Thr Gly Val  
 245 250 255  
 Ser Met Thr Asn Thr Leu His Ser Met Leu Lys Gly Leu Ala Pro Ala  
 260 265 270  
 Ala Ala Glu Ala Val Glu Thr Ala Ala Glu Asn Gly Val Trp Ala Met  
 275 280 285  
 Ser Ser Leu Gly Ser Glu Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu  
 290 295 300  
 Gly Ala Gly Val Ala Ala Asn Leu Gly Arg Ala Ala Ser Val Gly Ser  
 305 310 315 320  
 Leu Ser Val Pro Pro Ala Trp Ala Ala His Glu Ala Val Thr Pro  
 325 330 335  
 Ala Ala Arg Ala Leu Pro Leu Thr Ser Leu Thr Ser Ala Ala Glu Thr  
 340 345 350  
 Ala Pro Gly His Met Leu Gly  
 355

(2) INFORMATION FOR SEQ ID NO:165:

(a) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3577 base pairs

(B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

AGTTGAGTGC AGAATGATAC TGAGGGGCTG TATCCAGGAT GAGTGAGACA ACGGAAGGAC	60
CGTGGAGTTC GGGGACATCG CAAGCCGACG CAGTGGCTTT GGGCGGGAAG GCGGAGGGCG	120
CCGAAGCGGA AGAGCTGGCC GCGGGGAGCG GCGCGGTTTC CCGTGGCTTC CGGTGAAAGC	180
GTCAGGGGCT GAGATATGCC CCAGCGAGCG ACCAGAGAGT CCCCAGAGAT AGGAGAGCTG	240
GGAGAGCCCG GAGAGATATG ACGACTATCG CAGCTATGAG GCGGAGAGCC AGGAGGGCGC	300
ACGGTCGGCA TCTTGGCGAG GCGGCTTCCG GATTCGGTTA CCAAGACTGT CCAAGATTGC	360
CATCGGGGCG GCACTGTCTA TCACTCGGCG CTTCAGGAGC CTGAGCGGAT ACCTTGTGTG	420
GCAACAGCTT GAGGCGACCG AACCGGAGCG GCGGCGCGCG GCGTTCGCTG CCGGAGCGCA	480
GCAGAGTGTG ATCAACATCG CCGCTGTGCA CTCACAGAG GCCAAGAGAG AGGTGCGCGC	540
TGTCATCGAC AGCTCCACCG CGAATTGAG GATGAGCTTC CAGCAGCGGG CAGCGGATTT	600
CAGGAGGTTT GTGAAAGATT CGAAGTGTGT CACCGAGGCG AGGGTGAAGC CGGACGGCTT	660
CGAATCCATG AACAGGCAAT CCGGCTGTGT GCTGCTGCGC GCGACTTCAC GGTGACCGAA	720
TTGCGCTGGG GCGAAGAGAG AACCGCTGTC GTGCGGCGTC AAGTGTAGCG TTACGTAAGA	780
GCGGAGAGAG TACAGGATTT CGAAGCTTCA GTTCGTACCG TGAGGAGATG CGTACGCGAG	840
GTCACAGCGG AAACCACTCG GCGGACCGAA GTGCGTGAGA TGAGCTCAAC CCGAGAGGAA	900
GGCTGATGAT CCGGGGAGCG GGCATTGAGC AGCGACTCTG CAGCGGAATT TACGCTGCAG	960
AAGGTTTACC GCGAGCTGCA CCGTGTAGCA ATGCAAGTTA CCGTGAGACC CGTTCCGGTG	1020
ATTCTCATCC TCGCTCTGTT GATCTCTGAG GCGCGAGCGG GATGCTTATA CGTTGAGCGA	1080
TACGAGCGCA TCAAGAGAGC GACTGCGGCG CCGGCGGTCT TCGGCTGCGC GCGGCGCTCG	1140
AAGGAGCAAT CAGCTGTGTG TGTATTGAGC CGACAATTTC ACCAGAGCTT CCGTACCGCC	1200
AGGTGCGACC TCGCGCGCGA TTTCTGTCTC TATACGAGCA GTTCAGCGAG TAGATCGTGG	1260
CTTGGCGGCG CAAAGCGAGC TCACTGAAAG GCAACGCGCA GGTGTGCGCG GCGGCGGTGT	1320
CGGAGCTACA TCGGATTCG GCGGTGTCTC TGGTTTTTCT CAGGAGAGAG AGTACCAATTA	1380

AGGACAGCCG	CATCCCTCTG	ATGGGCGCCG	GCAGCGTGAAT	GGTGACCCGA	GCAGAGGTGG	1440
ACGGCAATTS	GCTGATCAAG	AGGTCACCC	CGGTTTAGGT	TGGCTATAGC	GGTCGCCAAG	1560
TCTGACGGGG	GGCGGGTGG	CTGCTGGTGC	GAGATACCGG	CGGTTCTCCG	GACAAATCAG	1560
GCCCGACCTG	AACACAGTCT	CGGCCCCGTG	CTAATCGGCC	GGGTTATTGA	AGATTAGTTG	1620
CGACTGTATT	TACCTGATGT	TGCAATGGT	CAGCTGGATT	TAGCTTCGGG	CGAGGGGGGG	1680
TGGTGACATT	TGCAATGCGG	GTTTGAGCTA	GTTGAGAGAA	TTTGACCTGT	TGGCGAGGTT	1740
GTTTGCTCTG	CATCATTCCT	GCTAGTTATG	GGGAGCGGGA	AGGATTATCG	AGTCGCTGGA	1800
CTTCGGGGGG	TTACCAAGCG	AGATCAACTC	CGGAGGATG	TACGCCGGCC	CGGTTTGGGC	1860
CTCGTGGTGG	GGCGCCCGGA	AGATCTGGGA	CAGCGTGGCG	AGTGAGCTGT	TTTCGCCGCC	1920
CTCGCGGTTT	CAGTCCGTGG	TGTGGGCTCT	GAGGAGCGGA	TGCTGATAG	GTTGCTCGGC	1980
GGGCTGATG	GTGGCGCGGG	CTTGCGGCTA	TGTGGCGTGG	ATGAGCTGCA	CCGCGGGGCA	2040
GGCGAGGCTG	ACGCGCCGCG	AGGTCGGGAT	TGGTGGCGCG	GCTTACGAGA	CGGCGGATGG	2100
GCTGAGGATG	CGGCGCGCGG	TGATGGCGGA	GAGCTGCTGT	GAGCTGATGA	TTGCTATAGC	2160
GAGCAACCTC	TTGGGCGAAA	ACAGCGCGGC	GATCGCGGTC	AGGAGGCGCG	AATACGGGGA	2220
GATGCGGGCG	CAAGACCGGG	CGCGATGCTT	TAGCTAGCGC	GGCAGGGGCG	CGAGCGCGAC	2280
CGAGGCGCTG	CTGCCCTTGG	AGGACGCGCG	AGTATGCTCT	AAGCGCGCGG	GCTGCTTGA	2340
GCAGGCTCTC	GCGGTCGAGG	AGGCGATGGA	CAGCGCGCGG	GCGAAGCTAGT	TGATGAGGAA	2400
TATGCGCGAA	CGGCTGCAAG	AGCTGGCGGA	AGCCACGAAA	AGCATCTGCG	GCTTCCGCGA	2460
AGTAGTGAAG	CTCTGAGAGG	CGATCTCGCG	GCACTCTGCG	CGGCTCGAGG	AGATCTGCTC	2520
GATGCTCAAC	AAGACGCTGT	CGATGAGGAA	CTCGGCTGTT	TGATGCGGCA	GAGCTTTCGA	2580
CTCATGCTTG	AAGGCGCTTG	CTCTGCGCGG	GCTTCAGGCG	GTGGAAGGCG	CGGCGGAGAA	2640
CGGCTCTGAG	GCGATGAGCT	GAGTGGCGAG	CGAGCTGCTT	TGCTGCTGCG	GTTCTTTCGGG	2700
TGTGCGGCTT	GGGCTGCTCG	CGAAGTTGGG	TGCGGCGGCT	TGCTGCTGCT	GTTTGTGCGT	2760
GGCGGAGGCT	TGGGCGCGCG	CGAAGCGGCG	GCTCACCGCG	GCGCGCGGCG	GCTGCGGCTT	2820
GAGTGGGCTG	ACGAGCGCGG	CGAAGCGGCG	CGGCGGAGCG	ATGCTGCGCG	GCTTACGCTT	2880
GAGTGGGCTG	ACGAGCGGCG	CGAAGCGGCG	CGGCGGAGCG	ATGCTGCGCG	GCTTACGCTT	2940
GCGCGGCGCG	TAGGTAATGC	CTCGTGTGCG	CGGCGGCGCG	TAGGCGCGAT	CGGCGGCGAA	3000

TCGACGGGCT CTATGCGGGC AGCGATC

3521

(2) INFORMATION FOR SEQ ID NO:106:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 396 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:106:

Val	Val	Asp	Phe	Gly	Ala	Leu	Pro	Pro	Glu	Ile	Asn	Ser	Ala	Arg	Met
1				5					10					15	
Tyr	Ala	Gly	Pro	Gly	Ser	Ala	Ser	Leu	Val	Ala	Ala	Ala	Lys	Met	Trp
		20						25					30		
Asp	Ser	Val	Ala	Ser	Asp	Leu	Phe	Ser	Ala	Ala	Ser	Ala	Phe	Gln	Ser
		35					40					45			
Val	Val	Trp	Gly	Leu	Thr	Phe	Gly	Ser	Trp	Ile	Gly	Ser	Ser	Ala	Gly
		50				55					60				
Leu	Met	Val	Ala	Ala	Ala	Ser	Pro	Tyr	Val	Ala	Trp	Met	Ser	Val	Thr
		65				70				75				80	
Ala	Gly	Gln	Ala	Glu	Leu	Thr	Ala	Ala	Gln	Val	Arg	Val	Ala	Ala	Ala
			85						90					95	
Ala	Tyr	Glu	Thr	Ala	Tyr	Gly	Leu	Thr	Val	Pro	Pro	Pro	Val	Ile	Ala
			100					105					110		
Glu	Asn	Arg	Ala	Glu	Leu	Met	Ile	Leu	Ile	Ala	Thr	Asn	Leu	Leu	Gly
		115					120					125			
Gln	Asn	Thr	Pro	Ala	Ile	Ala	Val	Asn	Gln	Ala	Glu	Tyr	Gly	Gln	Met
		130				135						140			
Trp	Ala	Gln	Asp	Ala	Ala	Ala	Met	Phe	Gly	Tyr	Ala	Ala	Thr	Ala	Ala
		145				150				155				160	
Thr	Ala	Thr	Glu	Ala	Leu	Leu	Pro	Phe	Gln	Asp	Ala	Pro	Leu	Ile	Thr
			165					170					175		
Asn	Pro	Gly	Gly	Leu	Leu	Glu	Gln	Ala	Val	Ala	Val	Glu	Glu	Ala	Ile
			180					185					190		
Asp	Thr	Ala	Ala	Ala	Asn	Gln	Leu	Met	Asn	Asn	Val	Pro	Gln	Ala	Leu
		195					200					205			

Gln Gln Leu Ala Gln Pro Thr Lys Ser Ile Trp Pro Phe Asp Gln Leu  
 210 215 220  
 Ser Gln Leu Trp Lys Ala Ile Ser Pro His Leu Ser Pro Leu Ser Asn  
 225 230 235 240  
 Ile Val Ser Met Leu Asn Asn His Val Ser Met Thr Asn Ser Gly Val  
 245 250 255  
 Ser Met Ala Ser Thr Leu His Ser Met Leu Lys Gly Phe Ala Pro Ala  
 260 265 270  
 Ala Ala Gln Ala Val Glu Thr Ala Ala Gln Asn Gly Val Gln Ala Met  
 275 280 285  
 Ser Ser Leu Gly Ser His Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu  
 290 295 300  
 Gly Ala Gly Val Ala Ala Asn Leu Gly Arg Ala Ala Ser Val Gly Ser  
 305 310 315 320  
 Leu Ser Val Pro Gln Ala Trp Ala Ala Asn Gln Ala Val Thr Pro  
 325 330 335  
 Ala Ala Arg Ala Leu Pro Leu Thr Ser Leu Thr Ser Ala Ala Gln Thr  
 340 345 350  
 Ala Pro Gly His Met Leu Gly Gly Leu Pro Leu Gly Gln Leu Thr Asn  
 355 360 365  
 Ser Gly Gly Gly Phe Gly Gly Val Ser Asn His Leu Arg Met Pro Pro  
 370 375 380  
 Arg Ala Tyr Val Met Pro Arg Val Pro Ala Ala Gly  
 385 390 395

## (2) INFORMATION FOR SEQ ID NO:107:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1614 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

CATCGAGGG AGTGATGACC ATGCTGTGGC AGCAATGGCC AGCGAGATTA ATGCCGACGC	50
GCTGATGGCC GAGGAGGGTC CAGCTGCAGT GTTGGAGGCC GCGCGAGAGT GCGAGAGGCT	100
TTTGAGGCT CTGAGAGTCC AGGCTGTGCA GTTACGCGCC AGCTGAGACT CTCTGGGAGA	150

AGCCTGAGCT	GGAGGTGGCA	GCGACAGAC	GTTCCGGCT	GCACGCCG	TGGTGGCTG	240
GCTACAAAC	GGTCAACAC	AGGTCAGAG	GGTCCGATG	CAGGAGCGG	CGCAAGCTG	300
GGCATACAG	CAGGCGATG	CGACAGCG	GTCGTGGCG	GAGATGCCG	CGAACCAT	360
CAGTCAGGC	GTCTTACGG	CGACAACTT	CTTGGTATG	AACAGATCG	CGATGCGTT	420
GAGCGAGAT	GATTATTCA	TCCGTATGT	GAACAGCGA	GCCCTGGCAA	TGGAGGTCTA	480
CGAGGCGAG	ACCGCGTTA	AGACGTTT	CGAGAGCTC	GAGCGATGG	CTTGGATCT	540
TGATCCGCG	CGAGCGAG	CGACAGCA	CGGATCTTC	GGATGCGCT	CGCTGGCAG	600
CTGACAGCG	GTTCGCGAT	TGCGCGGG	GCTGCGGAG	ACCTCGGCG	AATGAGTGA	660
GATGAGCGG	CGGTGCGG	AGGTGAGCA	GCTGCGGAG	CAGGTGAGG	CTTGTTCAG	720
CGAGGTGCG	CGAGCGCG	CGGCAACCG	AGCGAGGAG	GAGCGCGCG	AGATGCGCT	780
GTCGCGCG	AGTGGCGGT	CGAGCGGCG	GTCGCGGT	CGATGAGCG	CGAGCGCGG	840
CGCGCGCG	CGCGCGCG	AGTGGCGGT	TGCGCGGAG	GAGTGGTGA	CGCGCGCGG	900
GTCGCGGT	CGAGCGGT	AGAGCGGT	TGCGCGGT	GTCGCGGT	CGCGCGGT	960
CGAGCGGT	CGAGCGGT	CGCGCGGT	GTCGCGGT	CGAGCGGT	CGAGCGGT	1020
CGAGCGGT	CGAGCGGT	CGCGCGGT	GTCGCGGT	CGAGCGGT	CGAGCGGT	1080
CGAGCGGT	CGAGCGGT	CGCGCGGT	GTCGCGGT	CGAGCGGT	CGAGCGGT	1140
CGAGCGGT	CGAGCGGT	CGCGCGGT	GTCGCGGT	CGAGCGGT	CGAGCGGT	1200
CGAGCGGT	CGAGCGGT	CGCGCGGT	GTCGCGGT	CGAGCGGT	CGAGCGGT	1260
CGAGCGGT	CGAGCGGT	CGCGCGGT	GTCGCGGT	CGAGCGGT	CGAGCGGT	1320
CGAGCGGT	CGAGCGGT	CGCGCGGT	GTCGCGGT	CGAGCGGT	CGAGCGGT	1380
CGAGCGGT	CGAGCGGT	CGCGCGGT	GTCGCGGT	CGAGCGGT	CGAGCGGT	1440
CGAGCGGT	CGAGCGGT	CGCGCGGT	GTCGCGGT	CGAGCGGT	CGAGCGGT	1500
CGAGCGGT	CGAGCGGT	CGCGCGGT	GTCGCGGT	CGAGCGGT	CGAGCGGT	1560
CGAGCGGT	CGAGCGGT	CGCGCGGT	GTCGCGGT	CGAGCGGT	CGAGCGGT	1620
CGAGCGGT	CGAGCGGT	CGCGCGGT	GTCGCGGT	CGAGCGGT	CGAGCGGT	1680

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 432 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

CTAGTGATG GGACATGG CATTTCTGC AGTCTCACTG CTTCTGTGT TGACRTTTC	60
GCACGGCGG GKAACGAG CACTGGGGTC GAGGACGGG TGGCTGCCA TATGTCGGG	120
AGCTTCGTA CTTCTGTGC GCGGAGAG ETTCTGTAG TGGGCGCA TGACACCTC	180
TCAGATGCC CTCAACGTA TAAACGAG AAGAGCGAG ACCGACGAA GGTGCACTC	240
GCCGATGCC GTTTTCCTT ATTCTACCG AACTGGCGT TGGCTATGC GAACATCCA	300
GTGAGCTGC CTTCGTGCA AGCATTGCG TGACCGGCT CAGTATGCT CGGCGCAGG	360
TTCTGAGCG GGTGTTCAG CTGGTAGCG GTGAGTCCG ATTTTCTG GACACCTAG	420
TACGCTCGG AA	432

(2) INFORMATION FOR SEQ ID NO:109:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 368 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

Met	Leu	Trp	His	Ala	Met	Pro	Glu	Asn	Thr	Ala	Arg	Leu	Met	1	5	10	15		
Ala	Gly	Ala	Gly	Pro	Ala	Pro	Met	Asn	Ala	Ala	Ala	Gly	Trp	Gln	20	25	30	35	
Thr	Leu	Ser	Ala	Ala	Leu	Asp	Ala	Gln	Ala	Val	Glu	Leu	Thr	Ala	Arg	35	40	45	50
Leu	Asn	Ser	Leu	Gly	Gln	Ala	Trp	Thr	Gly	Gly	Gly	Ser	Asp	Lys	Ala	55	60	65	70
Leu	Ala	Ala	Ala	Thr	Pro	Met	Val	Val	Trp	Leu	Gln	Thr	Ala	Ser	Thr	75	80	85	90
Gln	Ala	Lys	Thr	Arg	Ala	Met	Gln	Ala	Thr	Ala	Gln	Ala	Ala	Ala	Tyr	95	100	105	110



Thr Glu Ala Met Ala Thr Thr Pro Ser Leu Pro Gly Ile Ala Ala Asn  
 100 105 110  
 His Ile Thr Glu Ala Val Leu Thr Ala Thr Asn Phe Phe Gly Ile Asn  
 115 120 125  
 Thr Ile Pro Ile Ala Leu Thr Glu Met Asp Tyr Phe Ile Arg Met Trp  
 130 135 140  
 Asn Glu Ala Ala Leu Ala Met Glu Val Tyr Glu Ala Glu Thr Ala Val  
 145 150 155 160  
 Asn Thr Leu Phe Glu Lys Leu Glu Pro Met Ala Ser Ile Leu Asp Pro  
 165 170 175  
 Gly Ala Ser Glu Ser Thr Thr Asn Pro Ile Phe Gly Met Pro Ser Pro  
 180 185 190  
 Gly Ser Ser Thr Pro Val Gly Glu Leu Pro Pro Ala Ala Thr Glu Thr  
 195 200 205  
 Leu Gly Glu Leu Gly Glu Met Ser Gly Pro Met Glu Glu Leu Thr Glu  
 210 215 220  
 Pro Leu Glu Glu Val Thr Ser Leu Phe Ser Glu Val Gly Gly Thr Gly  
 225 230 235 240  
 Gly Gly Asn Pro Ala Asp Glu Glu Ala Ala Glu Met Gly Leu Leu Gly  
 245 250 255  
 Thr Ser Pro Leu Ser Asn His Pro Leu Ala Gly Gly Ser Gly Pro Ser  
 260 265 270  
 Ala Gly Ala Gly Leu Leu Arg Ala Glu Ser Leu Pro Gly Ala Gly Gly  
 275 280 285  
 Ser Leu Thr Arg Thr Pro Leu Met Ser Glu Leu Ile Glu Lys Pro Val  
 290 295 300  
 Ala Pro Ser Val Met Pro Ala Ala Ala Ala Gly Ser Ser Ala Thr Gly  
 305 310 315 320  
 Gly Ala Ala Pro Val Gly Ala Gly Ala Met Gly Glu Gly Ala Glu Ser  
 325 330 335  
 Gly Gly Ser Thr Arg Pro Gly Leu Val Ala Pro Ala Pro Leu Ala Glu  
 340 345 350  
 Glu Arg Glu Glu Asp Asp Glu Asp Asp Trp Asp Glu Glu Asp Asp Trp  
 355 360 365

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 100 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

```
Met Ala Glu Met Lys Thr Asp Ala Ala Thr Leu Ala Glu Glu Ala Gly
 1             5             10             15

Asn Phe Glu Arg Ile Ser Gly Asp Leu Lys Thr Glu Ile Asp Glu Val
 20             25             30

Glu Ser Thr Ala Gly Ser Leu Glu Gly Glu Trp Arg Gly Ala Ala Gly
 35             40             45

Thr Ala Ala Glu Ala Ala Val Val Arg Phe Glu Glu Ala Ala Asn Lys
 50             55             60

Gln Lys Glu Glu Leu Asp Glu Ile Ser Thr Asn Ile Arg Glu Ala Gly
 65             70             75             80

Val Glu Tyr Ser Arg Ala Asp Glu Glu Glu Glu Glu Ala Leu Ser Ser
 85             90             95

Gln Met Gly Phe
100
```

(12) INFORMATION FOR SEQ ID NO:111:

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 396 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

```
GATCTGGGCG CAGCTGAAA CCAGATGTA GAGATGGAG TGAGGCGAG GTTCTTGCA      60
GGCCCATCG CCGCGCGCG CGGAGACGC GCGGAGGCC GCGTGGTGC GCTTCGAAA      120
AGCAACCAAT AAGCAGAAC AGCACTCGA CGCATCTCG ACGATATTC CTGAGGCCG      180
CGTCCAAATC TCGAGGCGCG ACGAGAGCA GCAGCAGCG CTCTCTTCG AATGGGGTT      240
```

```

CTGACCCGCT AATACGAAA GAACCGGAGC AAAACATCA CAGAGCAGCA GTGGAATTC      350
GCGGATATCG AGCGCGGCGC AAGGCGAATC CAGGGAATC TCAGTCGAT TCATTCCCTC      360
CTTGACGAGG GCAAGCAGTC CCTGACGAGC CTGCGA      396

```

## (2) INFORMATION FOR SEQ ID NO:112:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 80 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: Linear

## (a) SEQUENCE DESCRIPTION: SEQ ID NO:112:

```

Ile Ser Gly Asp Leu Lys Thr Gln Ile Asp Gln Val Gln Ser Thr Ala
1             5             10             15

Gly Ser Leu Gln Gly Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala Gln
20             25             30

Ala Ala Val Val Arg Phe Gln Gln Ala Ala Asn Lys Gln Lys Gln Gln
35             40             45

Leu Asp Gln Ile Ser Thr Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser
50             55             60

Arg Ala Asp Gln Gln Gln Gln Gln Ala Leu Ser Ser Gln Met Gly Phe
65             70             75             80

```

## (2) INFORMATION FOR SEQ ID NO:113:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 387 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: Linear

## (a) SEQUENCE DESCRIPTION: SEQ ID NO:113:

```

GTGATGCGG ATCGGCTTT TCCTATTCT AGCGAATC GCGCTGCGC TATCGAACA      60
TCCGATGAC GTTGCTTCT GTCGAGCCA TTGCTGACC GCTTCGCG ATGCTCGCG      120
CGAGCTTTC CAGCCTTTC TTCACTCGG TATGCTGCG GTCGATTTT TCTGGAAC      180

```

```

CCTGGTACGC CTCGAAACCG CTACCGCCCC NGGCGCTGC GAGCTGGTC AGGACTGCT      240
TCCCTCTGTC AAGGAGGGGA TGAATGGAGC TGACATTTCC CTGATTTGCG GTTGCCCGCG      300
CCTCGATACG CCGGAAATTC CACTGCTGCT CCGTCATGTT TTTGCTCGGT TTCTTTTCGT      360
ATTAGCGGCT CAGAGGCCCA TTGCGA      387

```

## (2) INFORMATION FOR SEQ ID NO:114:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 372 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (a) SEQUENCE DESCRIPTION: SEQ ID NO:114:

```

CGGACAGAGC ATCTCGTTG GCGAACGCC GCTGGGAGC GCTCCGTCG GGGGGGAGC      60
TGGCGGCGCG ATCTCTGTC TCGCGAGAC CGCGCTGGA TCGATCGACC AGTTCTACG      120
TTCCGAGACT TTGTTTGGT ATCTGTGGA TAGCGGTGAG CGCGGGCGCC AGCTCGGAG      180
TCTCGGAGG CAGCGCGGCT CGGTGTTTC GCGCGGAGC CAGACGCTCT CGACGAGAGC      240
CGCGGGGCTT CCGCGATTGC CATCTTGCC CA      272

```

## (2) INFORMATION FOR SEQ ID NO:115:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 20 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

## (a) SEQUENCE DESCRIPTION: SEQ ID NO:115:

```

Asp Pro Val Asp Ala Val Val Asn Thr Thr Cys Asn Tyr Gly Gln Val
1           5           10           15

Val Ala Ala Leu
20

```

## (2) INFORMATION FOR SEQ ID NO:116:

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 15 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:116:

Ala	Val	Gln	Ser	Gly	Met	Leu	Ala	Leu	Gly	Thr	Pro	Ala	Pro	Ser
1				5					10					15

(2) INFORMATION FOR SEQ ID NO:117:

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 19 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:117:

Ala	Ala	Met	Lys	Pro	Arg	Thr	Gly	Asp	Gly	Pro	Leu	Gln	Ala	Ala	Lys
1			5					10							15

Gln Gly Arg

(2) INFORMATION FOR SEQ ID NO:118:

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 15 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:118:

Tyr	Lys	Trp	Gly	Pro	Gly	Gln	Pro	Phe	Asp	Pro	Ala	Trp	Gly	Pro
1			5					10						15

(2) INFORMATION FOR SEQ ID NO:119:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 amino acids  
(B) TYPE: amino acid  
(C) STRANDNESS:  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:119:

Asp Ile Gly Ser Glu Ser Thr Glu Asp Gln Glu Asn Ala Val  
1 9 10

12. INFORMATION FOR SEQ ID NO:120:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(x1) FREQUENCY DESCRIPTION: SEQ ID NO:120:

Ala Glu Glu Ser Ile Ser Thr Xaa Glu Xaa Ile Val Pro  
1 5 10

(21) INFORMATION FOR SEQ ID NO:21:

```
(L) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 17 amino acids
      (B) TYPE: amino acid
      (C) STRANDEDNESS:
      (D) TOPOLOGY: linear
```

[41] SEQUENCE DESCRIPTION: SEQ ID NO:121:

Asp Pro Glu Pro Ala Pro Pro Val Pro Thr Thr Ala Ala Ser Pro Pro  
1 5 10 15

348

(2) INFORMATION FOR SEQ ID NO:122:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 15 amino acids

150

(B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:122:

Ala	Pro	Lys	Thr	Tyr	Xaa	Glu	Gln	Leu	Lys	Gly	Thr	Asp	Thr	Gly
1				5				10					15	

(2) INFORMATION FOR SEQ ID NO:123:

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 30 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:123:

Asp	Pro	Ala	Ser	Ala	Pro	Asp	Val	Pro	Thr	Ala	Ala	Gln	Leu	Thr	Ser
1			5					10					15		
Leu	Leu	Asp	Ser	Leu	Ala	Asp	Pro	Asp	Val	Ser	Phe	Ala	Asp		
		20					25					30			

(2) INFORMATION FOR SEQ ID NO:124:

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 22 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:124:

Asp	Pro	Pro	Asp	Pro	His	Gln	Saa	Asp	Met	Thr	Lys	Gly	Tyr	Tyr	Pro
1			5					10					15		
Gly	Gly	Arg	Arg	Xaa	Phe										
		20													

(2) INFORMATION FOR SEQ ID NO:125:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:125:

Asp Pro Gly Tyr Thr Pro Gly  
1 5

(2) INFORMATION FOR SEQ ID NO:126:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: /note= "The Second Residue Can Be Either a Pro or Thr"

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:126:

Kaa Xaa Gly Phe Thr Gly Pro Gln Phe Tyr  
1 5 10

(2) INFORMATION FOR SEQ ID NO:127:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 9 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: /note= "The Third Residue Can Be Either a Gln or Leu"

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:127:

Xaa Pro Xaa Val Thr Ala Tyr Ala Gly  
1 5



## (2) INFORMATION FOR SEQ ID NO:128:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 9 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

Xaa Xaa Xaa Glu Lys Pro Phe Leu Arg  
 1 5

## (2) INFORMATION FOR SEQ ID NO:129:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 15 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

Xaa Asp Ser Glu Lys Ser Ala Thr Ile Lys Val Thr Asp Ala Ser  
 1 5 10 15

## (2) INFORMATION FOR SEQ ID NO:130:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 15 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

Ala Gly Asp Thr Xaa Ile Tyr Ile Val Gly Asn Leu Thr Ala Asp  
 1 5 10 15

## (2) INFORMATION FOR SEQ ID NO:131:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 15 amino acids

(E) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

Ala Pro Glu Ser Gly Ala Gly Leu Gly Gly Thr Val Gln Ala Gly  
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:132:

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 21 amino acids  
 (E) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

Xaa Tyr Ile Ala Tyr Xaa Thr Thr Ala Gly Ile Val Pro Gly Lys Ile  
 1 5 10 15

Asn Val His Leu Val  
 20

(2) INFORMATION FOR SEQ ID NO:133:

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 882 base pairs  
 (E) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

GGAAGGCTGT CTTGGCTTT GGGTATATCG GTTTCGGCTC CTTGGGCTG GCGGTGGCGG	66
TGACATTCGG ACCGCGCGG GCTCAGAAAG CGGTAGGCG AGACAGAAAC GCGGAGCGAG	126
GGAAATTCAT GCGTTTATG CCGAGCGAAG AGCAGGGGAC GTCCCGCGCG GTCGCGCGCG	186
ATGATCCGAC CTTGGGATTC CAGGGGCGGA CATTTCGGCG TATACAGAAC GTGATCCCGC	246

```

GGCCGCGTAC CTCACCCGAG GTGAGTGGGA GGCAGGTTT GCTTGGWCG GAGCGCCCG 300
CGCTGACCG TATTGTGGCT GGCAGGTCG CAGTCGGGT CCLGATCATC ATTCCGCGGT 360
TCCCGGTTG GCAAGCTGGA ATGCCAGCA TCCCGACCG ACGCGGAGC AGCGCGGTGA 420
CCACGTCGC GACGACGCG CCGACACCG GCGGACCGC GCGGTGACC ACGCGCCAA 480
CGACGCGCG GACCACCGG GTGACACCG CTCGACGAC GCGCGGAGC AGCGCGGTGA 540
CGACGCGCG AACGACGCT CCGCGACGA CGGTGCGCG GACGAGGTC GCTCGGACA 600
CGGTGCGCG GACGACGCT GCTCGGACA CGCGACGCG GACGAGGTC GCTCGGACA 660
CGACGCGCA GCGGACGCA CAACCAACCG AACGATGCG AACGAGGCG CAGACCGTGG 720
CGCGGACGA GGTGCGCGC GCTCGGACA GCGCGTGGC TGCGCGGAG GCGGAGGCG 780
GCGCGACTT ATTGCGCGC TTCTGATCG GCTCGGCGT TCCTGAGGT GCGAGGAGT 840
GCGCGTAT GCGGTGAGG TGGTGTGCG CTGTCTGAG GA 862

```

## (2) INFORMATION FOR SEQ ID NO:134:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 815 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:

```

CATCAACCA ACCGCTGCG GCGCGCGCG CCGCGGATC GCGCTGCGC CAGCGCGCG 60
CGGTGCTTC GGTGCGCGC TTGCGCGGT CAGCGCGCT GCGGCGAGC GCGTGGGTG 120
CTAGCGCGGT GTTACGCGC TCGTTGCGG GCGCGCGCG GCGACACCG GTACCGCGCA 180
TGGCGGCTT GCGCGCGCG GCGCGGTG CAGCGGTG CCGGTGCGA CGGTGCGCA 240
CGACGCGCG GCGCGGACG CGCGCGCGC GCGCGCGCG CAGCGCGCG GCGTGGGTG 300
TGTGCGCGT ACGCGCGCG CCGCGGTG CCGCGTACG GCGCGCGCA CTACGCGCG 360
ACCGCGCGG CGCGCGCGC CCGCGCGCG CCGCGTACG ACCGCGTCA CCGCGCGCG 420
GCGGTGCGG GATTGCGCA CTACCGCGC GACCGCGCG AAGTACTGC GGTACCGAG 480
CACTTCGCA CGACGCGCA GCGCGCGGT GCGCGCGCG TCGTGAAT GCGCGCGAT 540

```

AGCGGCTAGC	TGTGGGCTGC	GATCAACCTC	GATCATGATC	TCGAGGTBAC	GGTGACCCGC	600
CCCGCGGAAG	GAGGCGCTGA	ACTGGGGGTT	GAGCGGATCG	GCATCGGTT	GAGGCGATGC	660
CCAGGCCAAT	ACGGGGATAC	CGGGTGTGNA	AGCGCGCCGC	AGCGGAGGTT	CGGTTGGGCG	720
ACNGTGGTGC	GGTGGGCTG	TTACGGGCTT	GTGTCGAAC	ACGAGTAGCA	GGTGTGCTCC	780
GGGAGGGGCA	TCACCGCGC	GTTCGGTCAG	GTGCT			840

(2) INFORMATION FOR SEQ ID NO:135:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1152 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:135:

ACGAGCGGCT	GGTGAGGTC	TCGATGAGA	GATCTGCGG	ACTGCGGCG	GGGTTGAGC	60
CITCTCCAG	AGCAACTGCT	GAGATGCTC	CCCGCGGAAA	CAGGCGCTGA	TTTGACGCTC	120
TATGACCGT	TCGAGGAGA	GATCATCGCG	CAGATTGATA	TGGCAGCGGT	GGGCTAACAG	180
GTGGCGAGA	TGCTGCACT	GTATGTCTG	GACTCGGTGT	CAGGATCAG	CTTTCGGAC	240
GGCGGGCTGA	TGTTGTGAG	CGAGGAGTC	GGCGAGAGCC	ACTATCGGAT	CGAGAGGCTG	300
GGCGGATCA	GCTGTTTGG	GGGGCGAGC	ATGACAGGC	CTTTCATGCT	TGAGATGCTC	360
AGGGGTAGC	CGGACATCA	GCTTTTCAAG	ACGAGCGGC	ACTAGCAGCG	CGGATCTCA	420
ACAGCGGAG	TGTCAGGCG	GGGCGGCTC	CTTCAGCAAG	TTACCGCGAG	CGAGATGCTT	480
GGGTTCTGC	TGCTTTAAG	CGAGCGGATC	GTCTGAGGA	AGATCTGAA	TGAGAGGCGC	540
TTGATTCGG	CACACAGCT	GGGCAAGAC	ATTCTGAGA	GCATGCGCAG	GATGAGGAG	600
TGCTGGGCT	GGGTCGATG	ATCGGCTTC	CTGGGAGAT	TGAGGGGCTT	CGAGGAGAT	660
GGGCAAGAG	GTATCTTCA	CGGCTGAGG	GATCTGCTC	GGAGGCTT	CGATTTCCAG	720
GGGCGCTGA	CTGGCGGCG	GTTGACGCG	TTCACTGCA	TGCTAGGCT	CGGCTATTGC	780
CTGCTGTACA	AGAGATGCT	AGGCGGATC	GAGCTTACA	GGGAGAGCG	GTATATGCTT	840
TTGCTACAG	AGGATTTAG	AGGCAAGCA	AGCTCTGCT	CGGATTTGAG	CAGAGCTTC	900

GCTGAACCTC CTGGCCGGCT GCTCAGTGCC CTTACGTAAT CCGTGGGCC CAGGCCGGCC	980
CGCCGGCCGA ATACCGGCAG ATCGGACGC GATTGGCCG CCGCCGGTT GGAGCCGTGC	1020
ATACGCCCG CACACTCACC GCGAGGGAAC AGGCTTGCCA CCGTGGGCC GCGCGTGTCC	1060
GCGTCTACTT CGACACGCC CATCAGGTAG TGACACGTCC GCGCGACTTC CATTGGCTCC	1140
GTTGGGACG AG	1152

## (2) INFORMATION FOR SEQ ID NO:136:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 655 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

CTCTGCGCA TTGCGGAGG TTAGTTGCC GGTGGGTGTA GCGCCATGAG TGGGACGAC	60
CAGCAATGCG GCAACAGCAC GGAATCCGGT CAGGACGCC ACCTGGTCCA CCGGGCGGCT	120
CGCGTGAAGT CCGCGCTGCC CGGCTCTTTC CTTCGGGAGG GTCAACCGAC CTGTTTCGCG	180
CCTGGTTTGC CGCCATTATG CCGGCGGCCC GCGTCGGGCG GCGGATATCG CCGAANETCG	240
ATCAGGACAC CGAGATAGCG GGTCTGTGCA AGCTTTTGA GCGCGCGGCG GCGAGGTTTC	300
GCGGCGATT CTACTAGGCA GAGTGTGTC CCGATACGGA TCTGACCGAA CTCGCTGCGG	360
TGCAGGCAAC CTTCTTGGC GATGCGCGCG AGGATGCGCG CTGGACCGAT CTGTGCGCGC	420
TTGCGGAGCG GAGCGCGGTA GGTGGTCAAG TCGGTCCTAC GCTTGGGCTT TTGCGGAGCG	480
TCCCGACGCT GGTGCGGGTT GCGCGCGGAA AGCGCGGGGT CGGGTGCAT CAGGAATGCC	540
TACCGCGCG GCGACTGCAC GCGCGTGGC GCGCGGATGT CAGCGATTGG GAGATGATGC	600
TGCGGTTCAF ACTGCTGGAC CAGTGGGCGG AACGCTGCGA TTCCCGGACC GCGCA	655

## (2) INFORMATION FOR SEQ ID NO:137:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 167 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

```

Asn Ala Val Val Ala Phe Ala Val Ile Gly Phe Ala Ser Ileu Ala Val
1           5           10           15

Ala Val Ala Val Thr Ile Arg Pro Thr Ala Ala Ser Lys Pro Val Glu
20           25           30

Gly His Gln Asn Ala Gln Pro Gly Lys Phe Met Pro Leu Leu Pro Thr
35           40           45

Gln Gln Gln Ala Pro Val Pro Pro Pro Pro Asp Asp Pro Thr Ala
50           55           60

Gly Phe Gln Gly Gly Thr Ile Pro Ala Val Gln Asn Val Val Pro Arg
65           70           75           80

Pro Gly Thr Ser Pro Gly Val Gly Gly Thr Pro Ala Ser Pro Ala Pro
85           90           95

Glu Ala Pro Ala Val Pro Gly Val Val Pro Ala Pro Val Pro Ile Pro
100          105          110

Val Pro Ile Ile Ile Pro Pro Phe Pro Gly Trp Gln Pro Gly Met Pro
115          120          125

Thr Ile Pro Thr Ala Pro Pro Thr Thr Pro Val Thr Thr Ser Ala Thr
130          135          140

Thr Pro Pro Thr Thr Pro Pro Thr Thr Pro Val Thr Thr Pro Pro Thr
145          150          155          160

Thr Pro Pro Thr Thr Pro Val Thr Thr Pro Pro Thr Thr Pro Pro Thr
165          170          175

Thr Pro Val Thr Thr Pro Pro Thr Thr Val Ala Pro Thr Thr Val Ala
180          185          190

Pro Thr Thr Val Ala Pro Thr Thr Val Ala Pro Thr Thr Val Ala Pro
195          200          205

Ala Thr Ala Thr Pro Thr Thr Val Ala Pro Gln Pro Thr Gln Gln Pro
210          215          220

Thr Gln Gln Pro Thr Gln Gln Met Pro Thr Gln Gln Gln Thr Val Ala
225          230          235          240

Pro Gln Thr Val Ala Pro Ala Pro Gln Pro Pro Ser Gly Gly Arg Asn
245          250          255

```

Gly Ser Gly Gly Gly Asp Leu Phe Gly Gly Phe  
 260 265

## (12) INFORMATION FOR SEQ ID NO:138:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 174 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

## (xii) SEQUENCE DESCRIPTION: SEQ ID NO:138:

```

Ile Asn Gln Pro Leu Ala Phe Pro Ala Pro Pro Asp Pro Pro Ser Pro
1      5      10      15

Pro Arg Pro Pro Val Pro Pro Val Pro Pro Leu Pro Pro Ser Pro Pro
20      25      30

Ser Pro Pro Thr Gly Thr Val Pro Arg Ala Leu Leu Pro Pro Trp Leu
35      40      45

Ala Gly Thr Pro Pro Ala Pro Pro Val Pro Pro Met Ala Pro Leu Pro
50      55      60

Pro Ala Ala Pro Leu Pro Pro Leu Pro Pro Leu Pro Pro Leu Pro Thr
65      70      75      80

Ser His Pro Pro Arg Pro Pro Ala Pro Pro Ala Pro Pro Ala Pro Pro
85      90      95

Ala Cys Pro Phe Val Pro Val Pro Pro Ala Pro Pro Leu Pro Pro Ser
100     105     110

Pro Pro Thr Gln Leu Pro Ala Asp Ala Ala Cys Pro Pro Ala Pro Pro
115     120     125

Ala Pro Pro Leu Ala Pro Pro Ser Pro Pro Ala Gly Ser Ala Ala Ile
130     135     140

Arg Ala Leu Thr Gly Ala Thr Ser Ala Ser Thr Leu Gly His Arg Ala
145     150     155     160

Leu Pro Asp Asp Thr Thr Ala Arg Gly Cys Arg Arg Thr Gly
165     170

```

## (12) INFORMATION FOR SEQ ID NO:139:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 35 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(12) MOLECULE TYPE: peptide

(866) SPECIMEN DESCRIPTION: 100 TO 150000.

```

Gln Pro Pro Ala Glu Val Ser Asp Gln Arg Val Ser Gly Leu Thr Gly
1      5      10      15
Ala Val Gln Pro Ser Pro Arg Thr Thr Ala Glu Asp Pro Arg Pro Arg
20      25      30
Asn Arg Arg
35

```

(2) INFORMATION FOR SEC ID NO: 140:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 104 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(11) SEQUENCE DESCRIPTION: SEQ ID NO:140:

Arg	Ala	Asp	Ser	Ala	Gly	Cys	Thr	Cys	Arg	Trp	Cys	Xaa	Pro	Ala	Gln
1			5						10					15	
Cys	Arg	Arg	Pro	Ala	Met	Arg	Gln	Gln	His	Gly	Ser	Arg	Ser	Thr	Thr
		20					25						30		
Pro	Pro	Gly	Pro	Arg	Gly	Arg	Ser	Ala	Arg	Val	Arg	Pro	Gly	Arg	Leu
		35					40						45		
Phe	Pro	Trp	Ala	Gly	Ser	Ser	Asp	Val	Thr	Pro	Pro	Trp	Phe	Ala	Ala
	50					55					60				
Ile	Met	Pro	Ala	Arg	Arg	Val	Gly	Arg	Pro	Val	Trp	Pro	Xaa	Val	Asp
65					70				75					80	
Gln	Ala	Thr	Arg	Arg	Thr	Gly	Leu	Cys	Lys	Leu	Phe	Gln	Arg	Arg	Ala
				85				90						95	
Gly	Gln	Leu	Arg	Arg	Gln	Phe	Trp								



## 160

## (2) INFORMATION FOR SEQ ID NO:141:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 53 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (11) MOLECULE TYPE: other nucleic acid  
(A) DESCRIPTION: /desc = "PCR primer"

- (12) ORIGINAL SOURCE:  
(A) ORGANISM: *Mycobacterium tuberculosis*

## (11) SEQUENCE DESCRIPTION: SEQ ID NO:141:

GGATCATAT GGGCATCAT CATCATCATC ACGTCATCGA CATCATCGG ACC

53

## (2) INFORMATION FOR SEQ ID NO:142:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 42 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (11) MOLECULE TYPE: other nucleic acid  
(A) DESCRIPTION: /desc = "PCR Primer"

- (12) ORIGINAL SOURCE:  
(A) ORGANISM: *Mycobacterium tuberculosis*

## (11) SEQUENCE DESCRIPTION: SEQ ID NO:142:

CCTGATTCG GGCCTCGTT GCGCCGGCTT CATCTTGAC GA

43

## (2) INFORMATION FOR SEQ ID NO:143:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (11) MOLECULE TYPE: other nucleic acid  
(A) DESCRIPTION: /desc = "PCR Primer"

- (12) ORIGINAL SOURCE:  
(A) ORGANISM: *Mycobacterium tuberculosis*

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:  
GGATCTGCA GGTCTGAAG CACCGAGGG C 31
- (2) INFORMATION FOR SEQ ID NO:144:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid  
(A) DESCRIPTION: /desc = "PCR primer"
- (vi) ORIGINAL SOURCE:  
(A) ORGANISM: *Mycobacterium tuberculosis*
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:  
CTCTGAATC AGCCTCTGGA ATGCTGGGA T 31
- (2) INFORMATION FOR SEQ ID NO:145:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid  
(A) DESCRIPTION: /desc = "PCR primer"
- (vi) ORIGINAL SOURCE:  
(A) ORGANISM: *Mycobacterium tuberculosis*
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:  
GGATCCAGC CTGAGATGA CACCGATGC CTT 33
- (2) INFORMATION FOR SEQ ID NO:146:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (iii) MOLECULE TYPE: other nucleic acid  
(A) DESCRIPTION: /desc = "PCR primer"
- (vi) ORIGINAL SOURCE:  
(A) ORGANISM: *Mycobacterium tuberculosis*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

CGAGGATTC TCAGAGGACC ATTCCGAGG ACA

33

(x) INFORMATION FOR SEQ ID NO:147:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1993 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (vi) ORIGINAL SOURCE:  
(A) ORGANISM: *Mycobacterium tuberculosis*
- (ix) FEATURE:  
(A) NAME/KEY: CDS  
(B) LOCATION: 152..1273

(x) SEQUENCE DESCRIPTION: SEQ ID NO:147:

TTTCTTCTGA GGGGAGGCTG GTGGAGGAGG GGGGACACGA AGAGCTGTTC TGTTCACGGA	60
AGCATGCGCA AACCGCCGGA TAGCTGCGCG GACTGTGGGG GGACGTGAGG GAGCGCAAGC	120
CGCGAAATTC AAGAGCCAGG AAGGCTATGA C CTC AAA ATT CCT TTS CAT ACG	172
Val Lys Ile Arg Leu His Thr	
1 5	
CTG TTS GCG GTC TTS ACC GCT GCG CCG CTG CTR CTA GCA GCG GCG GCG	220
Leu Leu Ala Val Leu Thr Ala Ala Pro Leu Leu Leu Ala Ala Ala Gly	
10 15 20	
TGT GCT TCG AAA CCA CGG AGC GGT TCG CTT GAA ACG GTC GCG GAG GCG	266
Cys Gly Ser Lys Pro Pro Ser Gly Ser Pro Glu Thr Gly Ala Gly Ala	
25 30 35	
GCT ACT CTC GCG ACT ACC GCG GCG TCG TCG CCG ATG ACG TTG GCG GAC	316
Gly Thr Val Ala Thr Thr Pro Ala Ser Ser Pro Val Thr Leu Ala Glu	
40 45 50 55	
AGC GCT AGC ACG CTG CTC TAC CCG CTG TTC AAC CTG TGG GCT GCG GCG	364
Thr Gly Ser Thr Leu Leu Tyr Pro Leu Phe Asn Leu Trp Gly Pro Ala	
60 65 70	

TTC CAC CAG AGG TAT GCG AAC GTC ACG ATC ACC GCT CAG GGC ACC GGT Phe His Glu Arg Tyr Pro Asn Val Thr Ile Thr Ala Gln Gly Thr Gly 75 80 85	412
TCT GGT GGC GGG ATC GCG CAG GGC GCG GCG GGG ACC GTC AAC ATT GCG Ser Gly Ala Gly Ile Ala Gln Ala Ala Ala Gly Thr Val Asn Ile Gly 90 95 100	460
GCC TCC GAC GCC TAT CTC TCG GAG GGT GAT ATC GCG GCG CAC AAG GGG Ala Ser Asp Ala Tyr Leu Ser Glu Gly Asp Met Ala Ala His Lys Gly 105 110 115	508
CTG ATG AAC ATC GGG CTA GCC ATC TCC GCT CAG CAG GTT AAC TAC AAC Leu Met Asn Ile Ala Leu Ala Ile Ser Ala Gln Val Asn Tyr Asn 120 125 130 135	556
CTG CCC GGA GTG ACC GAG CAG CTC AAG CTC AAC GGA AAA GTC CTC GCG Leu Pro Gly Val Ser Glu His Leu Lys Leu Asn Gly Lys Val Leu Ala 140 145 150	604
GCC ATG TAC CAG GGC ACC ATC AAA ACC TCG GAC GAC GCG CAG ATC GCT Ala Met Tyr Gln Gly Thr Ile Lys Thr Trp Asp Asp Pro Gln Ile Ala 155 160 165	652
GCG CTC AAC GCG GGC GTG AAC CTG CCG GGC ACT GCG GTA GTT CCG CTC Ala Leu Asn Pro Gly Val Asn Leu Pro Gly Thr Ala Val Val Pro Leu 170 175 180	700
CAC GCG TCC GAC GGG TCC GGT GAC ACC TTC TTS TTC ACC CAG TAC CTC His Arg Ser Asp Gly Ser Gly Asp Thr Phe Leu Phe Thr Gln Tyr Leu 185 190 195	748
TCC AAG CAA GGT CCC GAG GCG TGG GAG AAG TCG GCG GGC TTC GAC ACC Ser Lys Gln Asp Pro Glu Gly Trp Gly Lys Ser Pro Gly Phe Gly Thr 200 205 210 215	796
ACC GTC GAC TTC CCG GCG GTC CCG GGT GCG CTC GGT CAG AAC GGC AAC Thr Val Asp Phe Pro Ala Val Pro Gly Ala Leu Gly Glu Asn Gly Asn 220 225 230	844
GCG GGC ATG GTG ACC GGT TCC GCG GAG ACA CCG GCG TCG GTC CCG TAT Gly Gly Met Val Thr Gly Cys Ala Gln Thr Pro Gly Cys Val Ala Tyr 235 240 245	892
ATC GCG ATC ACC TTC CTC GAC CAG GCG AAT CAA GGG GGA CTC GCG GAG Ile Gly Ile Ser Phe Leu Asp Gln Ala Ser Gln Arg Gly Leu Gly Glu 250 255 260	940
GCC CAA GTA GCG AAT AGC TCT GGC AAT TTC TTG TTG GCG GAC GCG CAA Ala Gln Leu Gly Asn Ser Ser Gly Asn Phe Leu Leu Pro Asp Ala Gln 265 270 275	988
AGC ATT CAG GCG GCG GCG GCT GCG TTC GCA TCG AAA ACC CCG GCG AAC Ser Ile Gln Ala Ala Ala Ala Gly Phe Ala Ser Lys Thr Pro Ala Asn 280 285 290	1036

280	285	290	295	1084
CAG GCG ATT TCG ATG ATC GAC GGG GCG GCC CCG GAG GCG TAC CCG ATC	Gln Ala Ile Ser Met Ile Asp Gly Pro Ala Pro Asp Gly Tyr Pro Ile	300	305	310
ATC AAC TAC GAG TAC GGC ATC CTC AAC AAC CGG CAA AAG GNC GCG GCG	Ile Asn Tyr Gln Tyr Ala Ile Val Asn Asn Arg Gln Lys Asp Ala Ala	315	320	325
ACG GCG CAG ACC TTC GAG GCA TTT CTG CAC TGG GCG ATC ACC GAG GCG	Thr Ala Gln Thr Leu Gln Ala Phe Leu His Tsp Ala Ile Thr Asp Gly	330	335	340
AAC AAG GCC TCG TTC CTC GAG CAG GTT CAT TTC GAG CGC CTC CGG CGC	Asn Lys Ala Ser Phe Leu Asp Gln Val His Phe Gln Pro Leu Pro Pro	345	350	355
GCG CTC CTC AAG TTC TCT GAC CGC TTG ATC GCG ACG ATT TCG ACC	Ala Val Val Lys Leu Ser Asp Ala Leu Ile Ala Thr Ile Ser Ser	360	365	370
TAGCGTGTGT GAGCAGCAAG CGACAGCAAG CTGCGTCGGG CCATCGGGGT GCTTTGCGGA				1335
GCAFGCTHGG CGGTGCGGCT GAATGCGGC GCGTGGGCG CGCATCGCG TGTTCGGGTG				1395
GAATAGGTGC GTGTATCGCG CTGCTTGGCG TGTTCTTGCT GTGTGTGGTG CTGTCACTCG				1455
AGCGCATGGG TCGGATCAGG CTCACCGGCT TCGATTCTTT GACCGCGAGC GAATGGAATC				1515
CGCGCAAGAG CTACAGCGAA ACCGTTGTCA CGACAGGCTC GCGCATCGCG TCGGGCGCTA				1575
CTACGGGCGG TTGCGGCTCA TGTTGGGAGC GTTGGCGAGC TGCGCATCG CGCTGATCAT				1635
CGGGGTGCGG GTCTGTGTAG GAGCGCGGCT GGTGATCTTG GAACGGCTCG CGAAACGGTT				1695
GCGCGAGGCT GTGGGATAG TCGTGGGATT GCTCGCGCGA ATCGCGACGC TGTGTCTGCG				1755
TCTTGGGCGG GCAATGAGCG TGCGCGGCTT CATCGCTCAT CACATGCTTC CGCTGATCGC				1815
TACACAGCGT CGCATGTCGC CGGTGCTGAA CTACTTGCAC CGCGACCTCG GCACCGGGGA				1875
GGGCAIGTTC CTCTCCGCTC TGTCTTGAGC GGTGATGCTC GTTCCCATTA TCGGACCAAC				1935
CATCATGAGC GTTCTTGGCG AGGTGCGGCT GTTGGCGGCG GAGCGCGGGA TCGGGAATTC				1995

(2) INFORMATION FOR SEQ ID NO:148:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 374 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:145:

```

Val Lys Ile Arg Leu His Thr Leu Leu Ala Val Leu Thr Ala Ala Pro
  1           5           10           15
Leu Leu Leu Ala Ala Ala Gly Cys Gly Ser Lys Pro Pro Ser Gly Ser
      20           25           30
Pro Glu Thr Gly Ala Gly Ala Gly Thr Val Ala Thr Thr Pro Ala Ser
      35           40           45
Ser Pro Val Thr Leu Ala Glu Thr Gly Ser Thr Leu Leu Tyr Pro Leu
      50           55           60
Phe Asn Leu Trp Gly Pro Ala Phe His Glu Arg Tyr Pro Asn Val Thr
      65           70           75           80
Ile Thr Ala Gln Gly Thr Gly Ser Gly Ala Gly Ile Ala Gln Ala Ala
      85           90           95
Ala Gly Thr Val Asn Ile Gly Ala Ser Asp Ala Tyr Leu Ser Glu Gly
      100           105           110
Asp Met Ala Ala His Lys Gly Leu Met Asn Ile Ala Leu Ala Ile Ser
      115           120           125
Ala Gln Gln Val Asn Tyr Asn Leu Pro Gly Val Ser Glu His Leu Lys
      130           135           140
Leu Asn Gly Lys Val Leu Ala Ala Met Tyr Gln Gly Thr Ile Lys Thr
      145           150           155
Trp Asp Asp Pro Gln Ile Ala Ala Leu Asn Pro Gly Val Asn Leu Pro
      160           165           170           175
Gly Thr His Val Val Pro Leu His Arg Ser Asp Gly Ser Gly Asp Thr
      180           185           190
Phe Leu Phe Thr Gln Tyr Leu Ser Lys Glu Asp Pro Glu Gly Trp Gly
      195           200           205
Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val Pro Gly
      210           215           220
Ala Leu Gly Glu Asn Gly Asn Gly Gly Met Val Thr Gly Cys Ala Glu
      225           230           235           240
Thr Pro Gly Cys Val Ala Tyr Ile Gly Ile Ser Phe Leu Asp Gln Ala
      245           250           255
Ser Gln Arg Gly Leu Gly Glu Ala Gln Leu Gly Asn Ser Ser Gly Asn
      260           265           270

```

Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala Gly Phe  
 275 280 285  
 Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp Gly Pro  
 290 295 300  
 Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile Val Asn  
 305 310 315 320  
 Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala Phe Leu  
 325 330 335  
 His Trp Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asp Gln Val  
 340 345 350  
 His Phe Gln Pro Leu Pro Pro Ala Val Val Lys Leu Ser Asp Ala Leu  
 355 360 365  
 Ile Ala Thr Ile Ser Ser  
 370

## (ii) INFORMATION FOR SEQ ID NO:149:

## (i) SEQUENCE CHARACTERISTICS:

- (a) LENGTH: 1883 base pairs
- (b) TYPE: nucleic acid
- (c) STRANDEDNESS: single
- (d) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:149:

TGTTTTCGA CGCGAGBETH GTGGAGGARG GAGCUACCGA ACAGTCTCTC TCTTCGCLLA	60
AGCATGGGGA AACTCCTCGA TACTGTGGCG GCTGTCTCGG GACGTCTCAAG GAGCCCAAGC	120
GGCGAATTG AAGAGCCAGC AAGGCTATCG CTGGAATTT CTTTTCATA CGCTGTTGGC	180
CGTGTGAGC GTTGGGGGCG TGCTGCTAGC AGCCGCGGCG TTTGGCTCGA AACGAGGAG	240
CGTTCGGCT GAAACGGGCG GGGGGGCGCG TACTGTGGCG ACTACCGGCG GGTGCTGGCG	300
GGTGACCTG CGGAGAGCC GTACAGGCT GCTGTAGCG CTGTTCAAG TGTGGGCTGC	360
GGCTTTTAC CAGAGTATC GAGACTGAG GATCAGCGCT GAGGACAGCG GTTCTGGTGC	420
CGGATCGCG CAGCGCGCG GGGAGAGCT CAGATTTGCG GGTTCGAGCG CTATCTGCTC	480
GGAGGTGAT ATGGGCGCG ACAGGGGCT GTGACATG GCGTAGCGA TTTCCGCTCA	540
GCAGTCAAC TACAAGTGC CGGAGTAGG CGAGGAGTTC AGGTGAAGG GAAAGTCTCT	600

GCGGCGCTG	TACGAGGCA	CGATCAAAAC	CTGGAGGAC	CCGAGATCG	CTGGCTCAA	660
GXXUGGCTG	AACCTGCCG	GCACCGGAT	AGTTCCGCTG	CACCGCTGCG	ACGAGTCCGG	720
TGACACCTC	TTGTTACCC	AGTACCTGC	CAGCAAGAT	CCCGAGGGCT	GGGGCAATC	780
GCCGCGCTC	GGGACGCGG	TGGCTTGG	GCGGTCGGG	GCTGCGCTG	GTGAGAACGG	840
CAACGCGCG	ATGGTGACG	GTGCGCGCA	GAGACCGCG	TGGGTCGCT	ATATCGGAT	900
CGGCTTCTC	GAGAGGCA	GTCAACGGG	ACTCGCGAG	GCCCAATAG	GCAATAGCTC	960
TGGCAATTC	TTGTTCCCG	ACGGCGAAG	CATTCAAGG	ACCGCGGCTG	GCTTCGATC	1020
GAAACCGCG	GCGACGAGG	CGTTTGAT	GATCGACGG	GCGCGCGCG	AGGCTACCG	1080
GATCATCAN	TACGATAGG	CGTGTCAA	CAACCGCAA	AGGAGCGCG	CGCGCGGCA	1140
GAGCTTGAG	GCATTCTCG	CTGCGGAT	CGCGAGCGG	AACAAGGCT	GCTTCTGGA	1200
CGAGTTCTA	TTGCGCGCG	TGCGCGCGG	GCTGCTGAG	TGCTGCGG	GCTTATCGG	1260
GAGGATTTC	AGTAGGCTC	GTGACGAG	AGCGACGCG	AGCTCGGCT	GCGCGCTCG	1320
GCTGTTTGG	GAGGCTGCT	GCGGCTGCG	GCTGAGGCG	GCGCGCTG	GCGCGCTG	1380
CGGCGCTG	GTGAGGCG	TGCGGCTG	GCTGCTG	GCTGCTG	GCTGCTG	1440
GTGCTGCTG	TGCGGCTG	GCTGCTG	AGCTGAGG	GCTGCTG	GCTGCTG	1500
ACGGAATGA	ATCGAGGCA	CACTACGCG	GAGCGCTG	TGCGGCTG	GCTGCTG	1560
CGGCGCTG	CTACTAGG	GCTGCTG	TGCTGCTG	GAGCTGCG	AGCTGCTG	1620
TGCGGCTG	GATGCTG	GCTGCTG	TAGGAGGCG	GCTGCTG	GCTGCTG	1680
TGCGGAGG	GTGCGGCG	GCTGCTG	TAGGCTG	ATGCTGCG	GGAATCGCG	1740
GCTGCTG	GCTTCTG	GCGGCTG	GCTGCTG	GCTGCTG	GCTGCTG	1800
CTCGGCTG	GCTGCTG	GCTGCTG	TGCGGCTG	GCTGCTG	GCTGCTG	1860
GCGGCTG	GAGGCTG	TGCGGCTG	GCTGCTG	GCTGCTG	GCTGCTG	1920
TTATGCGCG	CAGGCTG	GCTGCTG	GCTGCTG	GCTGCTG	GCTGCTG	1980
CGATCGGGA	TTC					1993

(2) INFORMATION FOR SEQ ID NO:150:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 374 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear



[21] SEQUENCE DESCRIPTION: SEQ ID NO:150:

Met	Lys	Ile	Arg	Leu	His	Thr	Leu	Leu	Ala	Val	Leu	Thr	Ala	Ala	Pro	1	5	10	15
Leu	Leu	Leu	Ala	Ala	Ala	Gly	Cys	Gly	Ser	Lys	Pro	Pro	Ser	Gly	Ser	20	25	30	
Pro	Glu	Thr	Gly	Ala	Gly	Ala	Gly	Thr	Val	Ala	Thr	Thr	Pro	Ala	Ser	35	40	45	
Ser	Pro	Val	Thr	Leu	Ala	Glu	Thr	Gly	Ser	Thr	Leu	Leu	Tyr	Pro	Leu	50	55	60	
Thr	Asn	Leu	Tyr	Gly	Pro	Ala	Phe	His	Glu	Arg	Tyr	Pro	Asn	Val	Thr	65	70	75	80
Ile	Thr	Ala	Glu	Gly	Thr	Gly	Ser	Gly	Ala	Gly	Ile	Ala	Glu	Ala	Ala	85	90	95	
Ala	Gly	Thr	Val	Asn	Ile	Gly	Ala	Ser	Glu	Ala	Tyr	Leu	Ser	Glu	Gly	100	105	110	
Asp	Met	Ala	Ala	His	Lys	Gly	Leu	Met	Asn	Ile	Ala	Leu	Ala	Ile	Ser	115	120	125	
Ala	Glu	Glu	Val	Asn	Tyr	Asn	Leu	Pro	Gly	Val	Ser	Glu	His	Leu	Lys	130	135	140	
Leu	Asn	Gly	Lys	Val	Leu	Ala	Ala	Met	Tyr	Glu	Gly	Thr	Ile	Lys	Thr	145	150	155	160
Tyr	Asp	Asp	Pro	Glu	Ile	Ala	Ala	Leu	Asn	Pro	Gly	Val	Asn	Leu	Pro	165	170	175	
Gly	Thr	Ala	Val	Val	Pro	Leu	His	Arg	Ser	Asp	Gly	Ser	Gly	Asp	Thr	180	185	190	
Phe	Leu	Phe	Thr	Glu	Tyr	Leu	Ser	Lys	Glu	Asp	Pro	Glu	Gly	Tyr	Gly	195	200	205	
Lys	Ser	Pro	Gly	Phe	Gly	Thr	Thr	Val	Asp	Phe	Pro	Ala	Val	Pro	Gly	210	215	220	
His	Leu	Gly	Glu	Asn	Gly	Asn	Gly	Gly	Met	Val	Thr	Gly	Cys	Ala	Glu	225	230	235	240
Thr	Pro	Gly	Cys	Val	Ala	Tyr	Ile	Gly	Ile	Ser	Phe	Leu	Asp	Glu	Ala	245	250	255	

Ser Gln Arg Gly Leu Gly Gln Ala Gln Leu Gly Asn Ser Ser Gly Asn  
 263 265 270  
 Phe Leu Leu Phe Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala Gly Phe  
 275 280 285  
 Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp Gly Pro  
 290 295 300  
 Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Gln Tyr Ala Ile Val Asn  
 305 310 315 320  
 Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala Phe Leu  
 325 330 335  
 His Tyr Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asp Gln Val  
 340 345 350  
 His Phe Gln Pro Leu Phe Pro Ala Val Val Lys Leu Ser Asp Ala Leu  
 355 360 365  
 Ile Ala Thr Ile Ser Ser  
 370

## (D) INFORMATION FOR SEQ ID NO:151:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1777 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) SEQUENCE DESCRIPTION: SEQ ID NO:151:

GCTCTGAGC ACCACCTGGG TGTGGAATC GGTGACGGA TTGAAGTCA GGTACTCGT	60
GATGGGCGG GGAACACAT AGCACAAGC ATGCGAGCA GCGCGGTAC GGTGACCGT	120
GTAGCAAGC GCGAAGCGG CCGGCTTGG CACTTTTTC AGCGCTGTT TCACAAAGC	180
CCTCGAGAG GTGATCGCT GGAATTCTG CCGCGAAGC CTGCGACTA GCGCGAGCG	240
CTCGAAGCG GAGCGCGCG TGTGCAAGC GGTATCGCTT CAGCGGAGC GGTTCGCGG	300
GCGAAGCAT AGCATTTTC GACAAAGCT CTATACCTT GTGAGCGCT GCGCTACCG	360
AGCGCGGAC AGGTGTGAT ATGCTATTC TGCGCGACA GACAGGAGC AGGCTTTATG	420
ACAGATTTC GGTGAGAGC CTACGGGCG CGAAGTACC TAGAATGCG CGGAAAGCG	480
ATGGCTATA TCGACAGAG GAAAGTATC GGCATGCTT TTGAGAGCG CAGCGGAGC	540

```

TCCTTTACT TGTGGGCGA CATCATGCC CACTTGGAA GGTGGGCGG GCTGTGCC 600
TGGATCTTA TGGGATAGG CGCTCGGAC AACTCAGCC CATCGGACT GAGCGTAT 660
AGCTATGCG AGCAAGGGA GTTTTCTTC GCGTCTGGG ATGGCTCGA CTTGCGGAC 720
CAGCTGTAC TGTGCTGCA GACTGGGGC TGGGCTCGG GCTTGACTG GGTAAACCA 780
CATCGGAGC GATGCGAGG GATCGCTTC ATGGAAGCA TGTCTCGCC GATGAGCTG 840
GCGACTGCG CGCGGCGCT GCGGGGTGT TTCCAGGTT TCGATCGCC TCAAGCGAG 900
CGAATGGGT TGGAGCACA CATCTTTTC GAACGCTGC TGGCGGGCG GATCTCGGA 960
CAGCTCAGG ACGAGGAAT GAGCTCTAT CGGCGCTAT TGTGAACCG CGGCGAGAC 1020
CGTGGCCCA GTTTGCTTG GACAGAAC CTTCAGTCC ACGGTGAGC CGCGAGGTC 1080
GTGGCTTGG TCAACGATA CGGAGCTGG CTGAGGAAA GCGACATGC GAACTCTTC 1140
ACCAAGCGG AGCGGGCGG GATLATCAC GCGGCGATC GTGACTATG CAGGAGCTG 1200
CGCAACAGA CGGAATCAC AGTGGGCGC GTGCTTTC TTGAGGAGA CAGGATGCT 1260
GTGTATCTT GGGGGGGGG TCGCAGCAT GCGGAGCTC GAGCGCTCG CATTTACGA 1320
GACCAAGAT GTGATTTCC GCGAAGCGG GCGGCTGCT CTCACCTAT AAGCTTCT 1380
GTCGCGGCG GAGATCTCA GCGGAAGCG CAGCAATCG AGCGCTTCC TTGCAACGA 1440
GGTCGACAA TATAGCTGC AGGACAAGG TTTTCTATT TCGCAGCGA ATTAGTCT 1500
GCTTTCTAT GGGCTCAGT CAGGAAGCG GAGCGATCA GCGTATCG ATTGAGCTA 1560
TGAACCGGT ATCATGAAG CTTCAGTCA TGGAGACAG GGGGCTCTT GCGCGCTCG 1620
ATCTGCCAC AGGCTGCTC TCAAGCGC CACTGTACT ACGTTTACT AGCGGCCAG 1680
GCGATGCGG AGGAGCTCT GCGCGCTCG AGGAGCGAG GTATAGGCG GCTTTTCT 1740
TACTTGGCG TTGAGGTTG GCGCGCGCG GCTGCT

```

(2) INFORMATION FOR SEQ ID NO:152:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 324 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:152:

GAGATTGAAT	GGTACGGGTC	TCCTTAGGCG	CTCCGTCCCG	TGAATGCCCT	TATACGGCAC	60
GGCATGTTC	TGGGTGTGGA	CGTTGGGCCC	ATGCCCGGAC	GTGTGTAAAC	CGAGCGTTTG	120
ATCAGTAATT	CGGGGGGAG	GTTCGGGGA	GGCGGCGAC	ATGTTCGTGA	GGCGCGCGCC	180
CGCGTTGGCC	CAGGCGACCG	CTGGATGCTC	AGCGCGGTC	CGGCGAGTA	CGGAGCGTTT	240
GGCGGTGTC	CTCCACAGTA	CTACTCGGT	GACGACGGCG	CGCGGTGCTT	CGGTGAAGAC	300
CGTACCGAC	CGCGCGGATT	CAGA				324

## (x) INFORMATION FOR SEQ ID NO:153:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1335 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:153:

GGGTACCGC	CGGTTTGGC	TGGACGGGA	CGGTACGAC	CTGAACGACT	TGGCTCGCG	60
AAGATTGAC	GAGCGGTGG	TGGGGGGGT	GTGGTGTTC	AAGGTGTGG	CTGATCTGT	120
CGATGACCG	CGCGGACCC	GGGACTAGC	CTGTGAGAC	CTCTTGGCG	CGCGGAGGA	180
GCAGGATTC	CAGCGGACT	CGATCGCGT	GCTGACCGT	CGTGTGCTA	TGGGTGCTG	240
GGAGGTGCG	GTTCGGAAG	GATTTCGCT	CGTCACTGC	CTCGACGCG	AGGAGCAGCG	300
TTGGGCGCG	TGGACGAC	GGACGCGCG	CGAGTGGAG	AACCGGCTG	CGTGTCTGG	360
GTCTGATCA	ACTTGGCG	GATGTTGGG	TTCGCTGCG	AGGTTGCGG	CTGACGCGG	420
CTGAATCGAC	TAGATGAGG	CAGTTGGCA	CGATTCGGG	TGTGTGCGG	AGCAAGACAC	480
GATTCTGTC	ATCACTATT	GATGACTGC	ATGAGGCGC	TCATTGAGA	CGACCATGG	540
AAGTCCCGG	GGGAGAGCT	CTCGAGATC	ATCGGCGTC	CGTGGGAGC	CTGCGCTGT	600
GGCTCATTC	GGAGATGCT	CGGCTTGGG	GGATCTGCT	GACGCTAGC	CTGAAGGAGT	660
GGAGGCGCG	GGTCCGCGG	CTCTGCGG	GGCGGCGGC	GTGCTGCTG	CGTAAGGCGC	720
GGATGGCGA	GAAGCGCTC	GAGGTGGCG	CGCGGAGTT	CTTGTGCTG	CGGAGGCTG	780
GGGACAGCA	CGGAGAGCG	CTTGGCGCG	AGCGGCGGA	CTTGTGCGG	CGGAGGCGG	840

```

CCGACAGCAGC GAGCAGAGTGT GTGCTACTGT GGGCGGCAAC CAAAGTTCCT GGGGCACTGC      900
CGGTAAATCG GCGAGAGGGGT CTGGAGGGCA TCGAGGATCT GACGATCTGG ACCTGGAGT      960
AGGTGCGGCG GAGCGGGCTC GACTTTCGGC CCAAGACAA ACTGGGCTC TTGGTGGCTCT      1020
CGGCGATCCG GTTGGCCGAG CCGGTCCGGC TGGGGCTAG GCGGAGTAC GGGGTTCGA      1080
CGAGCTGGGT GCAAGTGGC CTCAGGGCA CTGTGGCGG GCGGCTGAC GACGAGGCG      1140
CGCTGGCGCA GTGAGCGCG GCGGTCCGGC AGGCGTGGC TTGACTGGC GCGATCGCT      1200
GGGTCTGAC TATACGCCA GTCGGCTG CAGTGAAT GTGTCTGTT CGGTCCCTGC      1260
TGGCTCAAT TGAGGCGCG GCGACAGCA GCAATGGCG GCGATCTCT GCGGCGCGG      1320
GGGCGGCGG CTACAGC

```

## (2) INFORMATION FOR SEQ ID NO:154:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 321 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:

```

GGGGGGGCG CCGGGGAGC GCGGTACG CCGGACCG GCTGAGCC GTGCTGTG      60
TGGGCTTGG GCGGAGCG GACCTTGGT TGGGTGGG GAGAGGGGT AAGGGGGAA      120
TAGCTGGCG GCGGTGCA GCGGGGTCC CCGGACCG GCGGCGCG GCGAAGGTG      180
GCACGGCGG TGGGGGCG GCGGCAAG ACCCGGAG CAGCGGCAAT CCGGCGCTA      240
AGGGGGGCA GCGGGGATC GCGGTGGCG GCGGGGCGG CGGCGGCG GCGGCGGCA      300
ACGCGGCGA TCGGCGAG C

```

## (2) INFORMATION FOR SEQ ID NO:155:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 192 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:

GAAGACCCGG	CCCCCCATA	TGCATCGGCT	CGCTGATAC	TTTCGCCAA	CGTCACGGG	60
GGGGCTGG	GCATCATC	ACCGTGCT	ACGGCCCA	CGGACCGG	TGGTGTGC	120
CGTGGGTC	CGAAGTC	ACTTCGGG	AAAGCGAG	GGAGGGCA	ATCAGCGG	180
CGGTCAGGA	TTGGGTGA	AAGAATTC	TGCAATCT	GCAGGGGA	CGTACGGT	240
ACGACCCAC	TGGGTGAG	CGTGGGCA	TAGGGGCT	GATACCGG	TTTCGCGG	300
GAGACTATC	GGTGGGG	GTCGAGGA	CGATGGGA	TTTCGGGG	TGGGGGAGT	360
TGGGGGGA	TGGGGTAC	GAGGGGTC	AATCATGG	CAGGAGGG	TATCTGCTA	420
ATGAGTTCT	GGGGGGGG	AGCAATAAG	GGAGGATC	GTGGGGGG	AGAGGGGA	480
ACCTCGGG	GT					492

## (12) INFORMATION FOR SEQ ID NO:156:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 536 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:156:

Phe	Ala	Gln	His	Leu	Val	Glu	Gly	Asp	Ala	Val	Glu	Leu	Tyr	Arg	Ala	1	5	10	15
Asn	Ala	His	Asp	Gln	Ala	Asp	Pro	Leu	Gln	Pro	Gly	Ser	Ala	Arg	Arg	20	25	30	
Gln	Arg	His	Ser	Arg	Ser	Pro	Arg	Arg	Leu	Ala	Gly	Pro	Asn	Ala	Tyr	35	40	45	
His	Tyr	Ser	Asn	Asn	Arg	Ser	Ile	Leu	Cys	Gln	Arg	Tyr	Pro	Leu	Pro	50	55	60	
Ser	Ala	Ala	Gln	Asp	Val	Ile	Cys	His	Leu	Cys	Pro	His	Arg	Gln	Glu	65	70	75	80
Pro	Gly	Leu	Met	Thr	Ala	Phe	Gly	Val	Glu	Pro	Tyr	Gly	Gln	Pro	Lys	85	90	95	
Tyr	Leu	Glu	Ile	Ala	Gly	Lys	Arg	Met	Ala	Tyr	Ile	Asp	Glu	Gly	Lys				

100					105					110						
Gly	Asp	Ala	Ile	Val	Pro	Glu	His	Gly	Asn	Pro	Thr	Ser	Sec	Tyr	Leu	
115					120					125						
Trp	Arg	Asn	Ile	Met	Pro	His	Leu	Glu	Gly	Leu	Gly	Arg	Leu	Val	Ala	
130					135					140						
Cys	Asp	Leu	Ile	Gly	Met	Gly	Ala	Ser	Asp	Lys	Leu	Sec	Pro	Ser	Gly	
145					150					155					160	
Pro	Asp	Arg	Tyr	Ser	Tyr	Gly	Glu	Glu	Arg	Asp	Phe	Leu	Phe	Ala	Leu	
165					170					175						
Trp	Asp	Ala	Leu	Asp	Leu	Gly	Asp	Ala	Val	Val	Leu	Val	Leu	His	Asp	
180					185					190						
Trp	Gly	Ser	Ala	Leu	Gly	Phe	Asp	Trp	Ala	Asn	Glu	His	Arg	Asp	Arg	
195					200					205						
Val	Glu	Gly	Ile	Ala	Phe	Met	Glu	Ala	Ile	Val	Thr	Pro	Met	Thr	Trp	
210					215					220						
Ala	Asp	Trp	Pro	Pro	Ala	Val	Arg	Gly	Val	Phe	Glu	Gly	Phe	Arg	Ser	
225					230					235					240	
Pro	Glu	Gly	Glu	Pro	Met	Ala	Leu	Glu	His	Asn	Ile	Phe	Val	Glu	Arg	
245					250					255						
Val	Leu	Pro	Gly	Ala	Ile	Leu	Arg	Glu	Leu	Ser	Asp	Glu	Glu	Met	Asn	
260					265					270						
His	Tyr	Arg	Arg	Pro	Phe	Val	Asn	Gly	Gly	Glu	Asp	Arg	Arg	Pro	Thr	
275					280					285						
Leu	Ser	Trp	Pro	Arg	Asn	Leu	Pro	Ile	Asp	Gly	Glu	Pro	Ala	Glu	Val	
290					295					300						
Val	Ala	Leu	Val	Asn	Glu	Tyr	Arg	Ser	Trp	Leu	Glu	Glu	Thr	Asp	Met	
305					310					315					320	
Pro	Lys	Leu	Phe	Ile	Asn	Ala	Glu	Pro	Gly	Ala	Ile	Ile	Thr	Gly	Arg	
325					330					335						
Ile	Arg	Asp	Tyr	Val	Arg	Ser	Trp	Pro	Asn	Glu	Thr	Glu	Ile	Thr	Val	
340					345					350						
Pro	Gly	Val	Sis	Phe	Val	Glu	Glu	Asp	Ser	Asp	Gly	Val	Val	Ser	Trp	
355					360					365						
Ala	Gly	Ala	Arg	Glu	His	Arg	Arg	Pro	Gly	Ser	Ala	Leu	Ile	Ser	Arg	
370					375					380						
Asp	Glu	Glu	Cys	Asp	Phe	Arg	Arg	Arg	Arg	Pro	Ala	Cys	Glu	Leu		
385					390					395					400	

```

Ile Arg Leu Pro Ala Pro Gly Arg Asp Ser Gln Gly Lys Gly His Gln
    405                                410                                415

Ser Gln Pro Leu Pro Ser Gln Arg Gly Arg Gln Ile Tyr Val Ala Gly
    420                                425                                430

Gln Arg Ser Ser Tyr Leu Pro Ser Gln Leu Val Ala Ala Phe Leu Trp
    435                                440                                445

Ala Gln Phe Gln Gln Ala Gln Arg Ile Thr Arg Ile Arg Leu Asp Leu
    450                                455                                460

Trp Asn Arg Tyr His Gln Ser Phe Gln Ser Leu Gln Gln Arg Gly Leu
    465                                470                                475                                480

Leu Arg Arg Pro Ile Ile Pro Gln Gly Cys Ser His Asn Ala His Met
    485                                490                                495

Tyr Tyr Val Leu Leu Ala Pro Ser Ala Asp Arg Gln Gln Val Leu Ala
    500                                505                                510

Arg Leu Thr Ser Gln Gly Ile Gly Ala Val Phe His Tyr Val Pro Leu
    515                                520                                525

His Asp Ser Pro Ala Gly Arg Arg
    530                                535

```

## (2) INFORMATION FOR SEQ ID NO:157:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 204 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (3) SEQUENCE DESCRIPTION: SEQ ID NO:157:

```

Asn Gln Ser Ala Pro Arg Ser Pro Met Leu Pro Ser Ala Arg Pro Arg
1      5      10      15

Tyr Asp Ala Ile Ala Val Leu Leu Asn Gln Met His Ala Gly His Cys
20     25     30

Asp Phe Gly Leu Val Gly Pro Ala Pro Asp Ile Val Thr Asp Ala Ala
35     40     45

Gly Asp Asp Arg Ala Gly Leu Gly Val Asp Gln Gln Phe Arg His Val
50     55     60

Gly Phe Leu Gln Pro Ala Pro Val Leu Val Asp Gln Arg Asp Asp Leu

```



65	70	75	80
Gly Gly Leu Thr Val Asp Trp Lys Val Ser Trp Pro Arg Gln Arg Gly	85	90	95
Ala Thr Val Leu Ala Ala Val His Glu Trp Pro Pro Ile Val Val His	100	105	110
Phe Leu Val Ala Gln Leu Ser Glu Asp Arg Pro Gly Gln His Pro Phe	115	120	125
Asp Lys Asp Val Val Leu Gln Arg His Trp Leu Ala Leu Arg Arg Ser	130	135	140
Glu Thr Leu Glu His Thr Pro His Gly Arg Arg Pro Val Arg Pro Arg	145	150	155
His Arg Gly Asp Asp Arg Phe His Glu Arg Asp Pro Leu His Ser Val	160	170	175
Ala Met Leu Val Ser Pro Val Glu Ala Glu Arg Arg Ala Pro Val Val	180	185	190
Gln His Glu Tyr His Val Val Ala Glu Val Glu Arg Ile Pro Glu Arg	195	200	205
Glu Gln Lys Val Ser Leu Leu Ala Ile Ala Ile Ala Val Gly Ser Arg	210	215	220
Trp Ala Glu Leu Val Arg Arg Ala His Pro Asp Gln Ile Ala Gly His	225	230	235
Gln Pro Ala Gln Pro Phe Gln Val Arg His Asp Val Ala Pro Gln Val	240	245	250
Arg Arg Arg Gly Val Ala Val Leu Lys Arg Asp Gly Val Thr Leu Ala	255	260	265
Phe Val Asp Ile Arg His Ala Leu Pro Gly Asp Phe	270	275	280

## (2) INFORMATION FOR SEQ ID NO:158:

- (1) SEQUENCE CHARACTERISTICS:
- (a) LENGTH: 284 base pairs
  - (b) TYPE: nucleic acid
  - (c) STRANDEDNESS: single
  - (d) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:158:

ATGAAATATCT	CGTCGGTGGT	GGGTGGCAAG	GCTTTGGGCT	GATCGGCGCG	CTATCTCTCC	60
GGCTGTGACG	CGATGCGGCG	TTTCTCGGAT	GGTTTGGGCT	AAGAGCTGCG	GGTTAGCGGA	120
ATCGGGGTCT	CGGTGATGCA	CGGAGCGCTG	ACCGAGACAC	CGCTGTGGCG	CAAGCTCGAC	180
CGGCGCGAGA	TGGGCGCGCG	GTTTGCGAGC	GTGAGGCGCA	TTCGGCTTCA	CTGGGTCGCG	240
GCAGCGGTCT	TTGAGCGGTCT	GGCG				264

(2) INFORMATION FOR SEQ ID NO:159:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1171 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:159:

TAGTGGGCGA	CGATGAGCTG	GCGGTCCAGG	CGGAGCGTT	CAAGCAACAG	CGGACCGCG	60
AAGCCGGTGG	GATGCTTACG	CGGGAAGCG	TGGGTGAGCA	CGGGCGCTCC	GGCGCAGAGC	120
AGTGTGAGCG	CAGCATATAG	CGCGGCTGCT	GCTTCAATTC	CGTTTGGGAA	TTCGGCGTAC	180
TGCTGGGTCA	TGTAGCGGCT	GGCGGCTTCA	TTTATCGACT	CGCTGGATTC	CGCGGAGTGG	240
CGGTTGAGCC	CGTCAATGGT	TAGAGGCTTC	TTGAGTGGCG	TTTCTTGGCG	CGCTGAGTCC	300
TGGGGTTCAT	GATCGGCGAG	GTGCGGGGAG	CGGAGCAGCT	GAGGCTCGAT	CGGTCGCGGA	360
ACCGGCTCTG	GACCGGCGCG	GGCAGCTTCC	CGGAGGAGCC	CGAGTGGCG	AGCTTGGGTT	420
ATCCCCAGCG	GGGCAAGCGT	TGGGCTGCT	CGGAGTTCG	GCAGAGGCT	GGCAGAGGAC	480
CGGCAATCAC	CAGGCAAGCG	TTGCGGAGTA	CGGATCTGCA	CTTGGCGATG	CGGCAAGCGA	540
ATCTGCTGCG	CGGCAATCTG	CAGATCGGCG	TGCTGGCTTG	ACAGGAGCGG	CGGCAATCTG	600
GGCAGCGGCT	ATCGGAGATG	GAGCGGCGCA	CGGATTTCTT	CAATCGCTCG	GGCTTGGCGG	660
ACTATTGCGA	CTTGGGCGCG	GTGCGGATAT	TGAGGAGCGA	TGGGAGTCTG	GAGGAGTCTG	720
CGGCAATCAA	CGGCAAGGCT	AGCTGCGGCG	GTGAGGCGCG	GAGTGGGCGG	GTGATTTTTC	780
CGGCAAGCTT	CGTGGGCTTT	GATGAGGCGG	TTGAGGCTTC	CGGCAAGGCG	GGGATCGAGC	840
TTATGAGCTT	CGGCAATCTG	CGGCAAGGCG	GTGAGGCGCT	TGCTGGAGCT	CAGGAGTCTG	900
ACATCGGCGA	CGGCAAGGCG	GTGAGGCGCG	GTGAGGCGCG	TGCTGGAGCT	CAGGAGTCTG	960

GAGTACAGCA ACATCGGGCC CTCAGCGTG TTCTCAACT ACCACGGTAC GTCCAAACGG	1020
GGCACCACCG TGGTTGGCA GGTACGGACC GGGTCGGACA CCACTTTCTT GGGCCAGATA	1080
ACCATCGGCG AGGGGGGTA TACCGGGGCC GGCACCTGG TGGGGAGGA TGTCCGCGCG	1140
GGGGCTGG CAGTGTGGC GGTCCGCA C	1171

## (2) INFORMATION FOR SEQ ID NO:160:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 227 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO:160:

GCAAGGCGG CAGGGGGGG GCGGGATGA ACAGCCGCA CCGCTGCTA GGGGCGAAG	60
AGGCGCGCA AGAGCGACC GCGACACCG GCGACACCG GCGCGCGCG GCGACACCT	120
TCACCCAGC CCGCAGCGC AACCGCGCA AGGCGGTGA GCGCGGGTC GCGCGAAGC	180
GCGGAAACG CCGAAAGGC GCGACACCA GCGACACCG CCGCGCG	227

## (2) INFORMATION FOR SEQ ID NO:161:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 304 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO:161:

CTCTGCACC ATGGGGGGG AGGGCGTGG CCGTGGGCG GGTCTACCG CAGCGCGAA	60
GGCGCGGCG GCTTCACTC CAGCGACCG GCGCGAGCG GCGGACGCG GCGACGAGG	120
CAACTGCCA GTGTCTGGG GCGAGCGCG CAGCGCGCG CATTGCGCA AGCGCGCGG	180
GGCGCGGCG GAGCGACCG GCGCGCGCG GCGCGAGCG GCTTTGGTG CGATGAGTC	240
CAGCGGCGC AATCTGGTG AAAAGCGCG AAGCGTAA CCGCGCGCA AGGTGGGCG	300

CGGC

386

## (2) INFORMATION FOR SEQ ID NO:162:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1439 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) SEQUENCE DESCRIPTION: SEQ ID NO:162:

GTGGGACGCT GCGGARKKTS TATAACAAAGS AUAACATGGA CCAGCGCGGG CTGGGTGAGC	60
TGATGACCTT ATTTAAAGAT GCGCGCTTCA GCGCGCGGCG GAGGACACCG GCGGAGGATC	120
TGATGCTGA GCTCTAGGAA TACTTCTTGG GCAATTTCGG TCGCGCGGAA GCGAAGCGGA	180
GTGGGACCTT CTTAACCCCG CCGGCGGTGG TCAAGGTGAT CGTGAGGTGG CTGAGCCGCT	240
CGAGTGGCGG GGTGATGAC CCGTGTGCGG GTTCCGAGGG CATGTTTGTG CAGACCCAGA	300
AGTTGATCTA CGAACAGGAC GCGGATCCGA AGGATGTCTC GATCTATGGC CAGGAAGGCA	360
TTGAGGAGAC CTGGCGGATG AGGAGATGTA AGCTGCGCAT CCGCGGATTC GACAACAAGG	420
GGCTGGCGGG CGGATGAGAT GATACCTTGG CCGCGGAGCA GCAACCGGAG CTGCAGATGG	480
ACTAGTGATG GCGCAATCGG CCGTTCACAC TCAAGGATCG GCGCGGCAAC GAGGAGAGCC	540
CAGCGTGGCG GTTGGGTGTT CCGCGCGGCG ATACGCGCAA CTGCGGATGG ATTGAGGACA	600
TCTTCTGCAA GTTGGCGCGG GAGGATGCGG CCGGCGGTGT GATGCGCAAC GGTTCGATGT	660
GCTGAGATTC CAGGCGAAGS GGGGATATTC GCGCGCAAT GGTGAGCGGG GATTGGTTT	720
CGTATGATGT GCGTTTACCG AGCTAGCTGT TCGGAGGAGC GGGGATCGCG GTGTGCGGT	780
GCTTTTTTGC CAAGAACAGS GCGGAGGTTA AGCAAGGCTC TATCAAGCGG TCGGCGGAGS	840
TGCTTTTCTT GAGGCTCTCT GACTGCGGCG AGCTATGAGA CCGGCTCGAG GCGCGCGTGA	900
CGAGAGAGGA GATGTTCCCG ATCGGAGATA CTTTCGAGCG GACGAGGAGC AGCGGCAAGC	960
CGGCTGCGGG TCGTGGCGGG GGTATGCGGG GCACTGCGCT CAAGGCGGCG GCGGCTGCTG	1020
GCGGCGGCGG CGGCAAGCGG GGTGCTGCGG GGTGCTGCTT CCGGAGGCTT GTGGGCGGCG	1080
AGCGGCGGGA GCGGCGGAGC GCGGCGGAGC GCGGCGGAGC GCGGCGGAGC GCGGCGGAGC	1140
GCAAGGCGGG CAAGGCGAGC AGCGGCGGCG GCACTGCGCT AGCGGCTCTT AGCGGCTCTT	1200

CGGGCCACGG	CGGCAACGGC	GCAATGGCG	GCAACGCGCG	CAACGGCTCG	GGGGGCGCG	1260
GGGGCCAGGG	CGGTGCGGCG	GGTACGGCG	GCAACGGCGG	CGACGGCGCG	CGTGGCAGCG	1320
GGGGGGCCAG	CGGCAAGGGG	GGCAACGGCA	CGAGGGCTGG	CGCGAGGGGG	TGAGGGCTCA	1380
TCAAGCTCAG	GGCGGGCGAC	GGCGGCAAGG	GGGCAATGG	CGGCAACGGG	GCAACGGCG	1440

(2) INFORMATION FOR SEQ ID NO:163:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 379 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:163:

GGGGGGGGCG	GGGGGATTT	TGTGTGGCT	TGATTGTGCG	TGGGGGTAAC	GGGGGTGATG	60
GTGTAACGCG	CGGATGGCG	GGGCTGGCG	GGGCTGGGCG	CGCGGGGGCG	GGCGGGGGCG	120
TGATCAGCT	CGTGGGGCG	CAAGGGCGCG	GGGGGGGGCG	CGGACGGCG	GGGGGGGGCG	180
GGTGTGGCG	TGCGGGGGCG	GGGGGGGGCG	CGGGGGGGCG	GGGCTTCAAC	GGAGGTGGCG	240
GGGGGGGGCG	GGGGGGGGCG	AGGCTGCTGG	GGGGGGGGCG	GGGGGGGGCG	GGGGGGGGCG	300
GGGGGGGGCG	GGGGGGGGCG	GGGGGGGGCG	GGGGGGGGCG	GGGGGGGGCG	GGGGGGGGCG	360

(2) INFORMATION FOR SEQ ID NO:164:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 68 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:164:

GGGACGGCTG	CAGGGGGCG	AGGAGGCGCG	CGGTGGGGAT	GGGGGGGGAT	GGGCTGCTG	60
GGGGGGGGAT	GGGAGGCGCG					80

(2) INFORMATION FOR SEQ ID NO:165:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 392 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO:165:

GGGCTGTGTG GCATCTGCAC CGCCGCTTC GCGACGTTG GCGGCCAAT ATCCAGCTCA	90
AGGCTACTA CTTACGTCG GAGGACGGC GCATCAGGT GCGGTGTC GCGCAAGGA	120
TCAAGGTCAT GCGGCGGAC GGGCATGAG GCGCTGCTG GCGGCTCGG GCGGATCGG	180
CGGGGGGCA CTTGCGGGC GAGCGGGT CATGCTGAG AACGCGGCG ATCTGTGAG	240
CACACTGAT GCGGCGGAC GAGATCTTC CAATTGTGA GCGGTGTTG ACCGCGGGA	300
CGGGTATAC GATATGCAAC CTATGCACT GCGGAGAAC GCGGTACGA TCGGTGATC	360
GCGGACAGC CCACGATGC AAGACCGTTA CA	392

(2) INFORMATION FOR SEQ ID NO:166:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 535 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:166:

ACCGGCGCA CGGCGGAC GGGGTTGCG GTGCGGCG GCGGCGCG GCGGCGGCG	60
GGTATGCGC GTGCGGCGC CACCAACGC TTGCTGCGC GTGCGGCGC GCGGCGGCA	120
GCGGCGGCG GCGGCTGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG	180
GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG	240
ACCGGCGCA CGGCGGCG GCGGCTGCG GTGCTGCG GCGGCGGCG GCGGCGGCG	300
GGTATGCGC GTGCTGCG GTGCTGCG GTGCTGCG GTGCTGCG GTGCTGCG	360
GGTATGCGC GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG	420
ACCGGCGCA CGGCGGCG GCGGCTGCG GCGGCGGCG GCGGCGGCG GCGGCGGCG	480

ACGGGAGGTC CCGGCGGGTC CCGGCGGAC AACGCGACG CTGCTGGTT CCGG 535

(2) INFORMATION FOR SEQ ID NO:167:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 699 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:167:

CGGAGCTGC CGGCGGATA CCGGCTGC CGGAGCTGC ATGCTGCGA CGGAGATCG	60
CGGCTGCTC CAGGCTGCT GCGGCTGCT GCTGCTGCT GCTGCTGCT GCGGCTGCT	120
CGGCGGAGC CCGGAGTGA CCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT	180
GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT	240
GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT	300
GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT	360
GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT	420
GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT	480
GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT	540
GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT	600
GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT	660
GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT GCGGCTGCT	699

(2) INFORMATION FOR SEQ ID NO:168:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:168:

ACGGTGAAGG CGGTACTGGC GCGGGGACG GCGGCAACGG CCGGAATCCG GGTGCGCTCT	60
TGGGACACAG CCGGGGATGG GCGACGGTG GCGCCGCGAG CACCGGTACT GCAAGTGCGG	120
GCTCTGGGGG CACCGGGGGG GACGGCGGGA GCGGCGGGGG TGGCGGGTTC TTATGCGGGG	180
CGGGCGCGGG GGGGCAAGGT GCGACTGGCG GCGGCGGGGG TCGCGGTGTC GACGGTGGGG	240
GCGCGCGGGG GAGCGGGGGG GCGGCGGCGA AGCGGCGGGC CCGGGGTCAA GCGCGCTGCG	300
TCTTGGGGGG CCGCGGCGAG GCGGAGCGGG GCGGCTAGGG GCGCGATGGC GGTGGGCGGG	360
GTGCGGGGCT CGACGGGCGG ATGCGCGGGC TGGGTGATGC CGGTGGC	407

(2) INFORMATION FOR SEQ ID NO:168:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 466 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(3) SEQUENCE DESCRIPTION: SEQ ID NO:169:

GATCGGTCAG CGCATGGGCG TCGGGGGCAA GCGATTGGCG GGTCTCCCGG AAGAACAATG	60
TCTCGGCGCG GCGCGGAGCG AGCGCGTGGC GCTGCGGCGC GTGGAAGGCG TCGAGAGGCG	120
ACAGCGAGTG CTGCGGCGCG TCGGAGGCGA AGAGTGGGCT GTACAGGGTG TAGATCGGCG	180
GGATCGCGCG CTGCGGCGAG GCATTGCGCG AGCGCGGCGG GTTTTGTGCA TCTCGGAGCA	240
TGACGCGCAT GCTCGGCGCG ACCAGGGGCG GCGCGGCGAA GGTGGCGCGG CTGGCGAGTA	300
GCGCGCGGAG GTGCGGCGCG AGGTGCTCGG GGAATGCGCG GCGGAGCGCT CCGCGCGGAG	360
GCGCGAAGAA CGATCGGCGA CCGGCGTGGG TCGGCGTGGC ATATCGGCGG GGTGCGTGGG	420
CGATATTGGA CCGGATGCGG CGACCGGCTT ACGGCGGCGG GAGGACCG	468

(2) INFORMATION FOR SEQ ID NO:170:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 219 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear



## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:170:

GCTGTAAAG GGGGCGAGG TGGCATGGG GGGGGGGGG AGAGAGGCG GAGCGGCGG	60
GGGCGCAATG CTACGGGGG AAACGGGGG AACGGGGTA GCGTGGTAA CGTGGGGAC	120
GGGGGGGGG GGGCAATG GGGGGGGG GCGAAGGCG AGCGGGGGG CTACGGGAC	180
GGGGGAGGG GCAAGGGGG GACGGGGGG AACGGGGG	240

## (2) INFORMATION FOR SEQ ID NO:171:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 494 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (x2) SEQUENCE DESCRIPTION: SEQ ID NO:171:

TAGCTCCGGC GAGGGGGCA AGGGGGGGA CGGTGGGAC GCGGTGACG GCGTGGGCG	60
CAACAGTTGC GTACGCAAG GCGGCAAGG CGGTGGGCG GCGGCGGCG GCGGCGGCG	120
GAGCGGTTT TTGCGGGCA AGGGCGGCT GGGGCGGAC GCGGTGAGG GCGGCGGCA	180
GCGGCGGCT AGGTGCGCA CGGTGGGCG TGGGGGGG AACGGGGTG TCGGCGGCG	240
GCGGCGGAC GCGGTCTTG GCGTGGGCG GCGGCGGCG GCGGTGCGT GCGGCGGCG	300
CAGTGGGCG GGTGCAAG GCGGCAAGG GCGGTGCG GCGGCGGCG GTGGGGGAG	360
CAACCGGCG GGTGCGCG AGTGGGCGT GCGGTGCG AACGCGGCG AGTGGGCG	420
CATCGGATG GCGGTGAGT GCGGCAAG CCGGCAAGT CAGTGGGCG AGGGGGGCG	480
ATTGCGGAT CAGC	494

## (2) INFORMATION FOR SEQ ID NO:172:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 220 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:172:

AGCGCGGTGG TGGGGGGGGC CAGCTCTTCA GGGCGGAGG CCGGGGGGT GCGTTGGGG	60
TTGGCGGAC CCGCGGCGAG GGTGGGGCTG GGGTGGCGG AGCGGGCGGC GCGGACGGCC	120
GCGGAGGAC AGGTGTATAC GGTGGTACCG GTTGGGTGG GCGGGCGGG GGGTGGGGC	180
GCGGAGGCTG CAGGGGCAAT GCGGGGGCA TGAACGGCTC	220

(2) INFORMATION FOR SEQ ID NO:173:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 330 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:174:

ATGGCGGCAA GGGGGGGGCG GGGGGGCTG GGGGGGCTG GACTACAT TGTGACGGC	60
GCGGAGGCTG GTGGGGGCG CCAAGGCGG CAAAGGCGG TGGGGGGAG AAGCACCAC	120
TGATGGGCTT AGCGGACCG GGGAAAGCG ATCCAAAGG GAGGATGGC GCGTTGGTTG	180
GCGGCTTGA CAGGGCGGG ATCACCTAC CTGAGCGAG CAGGGGATA AGGGCGGCA	240
AGGGATGTC TGGGCTGCT GCTAACGGG TACAGGCTT AGAGCTGGT GCGGAGCTGC	300
GGAATACAG TCGGGGCTG AGCATGGAC GGGGGGCA GTTCGCTGC ATCGCTCAC	360
GCGGCTAGC CCGGAGAC CTGAGCA	380

(2) INFORMATION FOR SEQ ID NO:174:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 400 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:174:

GCAAGGGCG CACCGGGGG GCGGGCTGA AGAGCTGCG GCGGCTGTA GCGGCCAAG	60
ATGGGGGCA AGGCGGAGC GAGGCGACG GGGGAGGC GGGGGGGC GGGGAGCTT	120

TGACGCACGG	CCCCGACGG	AACGCCGCG	AGACGGGTGA	CGGGGGGATC	GGCGGCACGG	180
GGGGAACGG	GGGAACGGC	GGACACACA	CGGCACCGC	GGCGCGGGC	AGACACGGC	240
GGACGGCGG	GGCGGGGGG	GGCGGCGGA	GGGGCGGAG	GGCGGGAGC	GGCGGGACG	300
GGCTGGCGG	CGACACAGC	AACGGCGCG	AGCGCGGAC	CGCGGGAAA	GGCGGCACG	360
GGGGGACGG	TGGATCTCA	GGCAGCAGC	GTCGTGGGG			400

## (2) INFORMATION FOR SEQ ID NO:175:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 538 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) SEQUENCE DESCRIPTION: SEQ ID NO:175:

GGCAACGGG	GGAGGGGGG	CATCGCGGG	ATTGGGGGG	AACGGCGTC	CGGACGGGG	60
AGCGGCACG	GGGGCACCG	GGGGACGGG	GGGACCGGG	GGCAAGGGG	GGATGGGGG	120
CGACGGGGG	ACCGCGGGG	GGGACGGGG	TGGGGGGGG	AACGGGGGG	CGGCGGGGG	180
GGGGGGGAC	GGGGGGGAC	GGGGGGGAC	GGGGGGGAC	GGGGGGGAC	GGGGGGGAC	240
GGGGGGGGT	GGGGGGGTA	GGGGGGGTA	GGGGGGGTA	GGGGGGGTA	GGGGGGGTA	300
GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	360
GGGGGGGGT	GGGGGGGGT	GGGGGGGGT	GGGGGGGGT	GGGGGGGGT	GGGGGGGGT	420
GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	480
GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	538

## (2) INFORMATION FOR SEQ ID NO:176:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 270 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear



CGTACTGCCC CGACACCTG GAACA

985

(2) INFORMATION FOR SEQ ID NO:178:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2136 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:178:

CGGCACGAG ATCGTATAC CGCGGATCG GAGCTGCGG ATTGCGCGGG TTTCGGCAGC	60
CGAGGAAGC CGTAGACGA TGCGCTGCG GAAGTAGCG CATCGTTGG CGATGCCGCG	120
ATGAACGGC GGCATCAAT TATGACGGA AGCTTTGAG TTACCGACGA TAATGGCTAT	180
ACCATTAAG AGGATGATCG GATATGAGC ATCGGAGAG CGTGACGGTG GATCAGCAG	240
AGATTITAA CAGGCGCAAC GAGGTGCGG CAGGATGCG GAGCGCCCG ACTGATGTCC	300
CGATCACAC GTGCGAAGTC AGGCGCGTA AAGCGCGCG CCACACAGTC GTATTGTCCG	360
CGGACAAAT CGGGAGTAC CTGGGAGCG GTGCCAAGA GCGGAGGCT CTGCGAGCT	420
CGTCCGGA CGCGCGCGAG CGGTATGCG AGCTTGATGA GAGGCTGCG AATCGGCTCG	480
ACAGCGAGG GAAAGAGCT GTGCGAGCG AATGCGCGG GCGGCTCGA CGGACAGTT	540
CGGCGAGCT AGCGATACG CGGAGGTCG CGCGGCGCG TGAGCCGAG TTGATGGATC	600
TGAAGAGAG GCGAGGAGG CTGAGAGCG GCGACCAAG CGATGCTTC GCGACATTG	660
CGGATGAGT GAACCTTTC AACTGAGCG TCGAGCGGA CTGAGAGCG TTCCGCGGCT	720
TTGACAGCT GAGAGCGGT GCGGTACCG CTGCGAGCG TTGCTGAT CAACAGCGCG	780
ATGCGACTT CGCATGCGC AATTGAGCG CTGCGATCG CAGCGAGCT CAATATGTCG	840
CGGAGCTGA CGTGTGCGT AGCGCGGAG ATCGAGCTT TGAAGACAT GTCGGCTCG	900
AAGCGCTTA CGCGGAGAG CTGTCGCGC GCGACGAAI TTTCGCGGTG TACCGGAGT	960
ATCAGCAGG GTGAGGAGG GTGCTGAGC AATACAGCA CAGGCGAGC CTGGAACCGG	1020
TGAACCGCG GAGGCTCGC CGCGCATCA AGATCGAGC GCGCGCGCT CGCAGAGAG	1080
AGGATTTAT AGCTGCTTC CTGATGCGC CGCTGACGG CTGCGGTGTC ACTGCGGTA	1140

```

CCGGGATGCC AGCGCGNDG ATGTTTCGCG CTACCGGATC GCGGGGTGAT GGCCTCCGG 1200
CTGACACGGC GCGCGAGCTG AGTCGGCTG GCGGGGAGCG CTACGCGCTG TCGGGGAGCG 1250
TGGCGGTCAA AGCGGATCG CTCGCTGCGG GTGGAGGGGG CGGGGTGCGG TCGGGGCGGT 1300
TGGGATCGCG GATCGGGGGC GCGGAATGGG TCGGGCGCGG TGGCGCTGGT GACATTCGGG 1350
GCTTAGGCTA GGAAGGGGGC GCGGGGGCGG CGGCGCTGGG GCGCGGTGGG ATGGGATTC 1400
CGATGGGTAC GCGCGATCGG GACAAAGGGG GCGGGAAGTC CAAGGTTTCT CAGCGGGAAG 1500
ACGAGGGGCT CTACACGGAG GATCGGGGAT GGAAGGAGCG GCTCATGGT AACCTCGGG 1550
GCCAGGACAG TAGGAGTGG AGTGAAGCAT GCACGATTG GACCGAGCTG TCGGCGGGG 1600
GTGAGGCTG GCGGGGGGT TTCAGTCGGG GGTAGAGGGG AGGTCGATC AGATGAACAA 1650
CGGATCTTC GCGCGGAGG AGGAAGCGCA GAGGTCGAA GTGAGGATCA ATGGGCGGCA 1700
GTGCTCGCC GCGCTGGGG TCGAGATGG TTTGCTGAG AGGCTGGATC CCGGGCGGT 1800
GGCTGAGGCG GTACAGAGAG GCTGGAAGA TGGGAGGGG GCGGGGTGGG GTATAGGGA 1850
CGGGGGGGG GAGCGAGTGA CGGTCGGGT ATCGGCGAG TCGGCGGGA TGAAGGAGG 1900
AATGGCTAA GCGCATTTT GCGGTGTTAG CAGTACGCA CGGATGAGC GCGGCAATGC 1950
GGCATTCAG CCGGGCGGAG AGGCGGTGAG TACGATTTGT CAGTGTTCG ACATGGATCG 2000
GCGGGTTCG GAGGGGGCGG TAGTCTGCT GCGCAATAT GCGGAGCTA GGTGGTCTTA 2100
GGTTCGGTGA GGTGGTGA TTATGAGTC GTTACCA 2150

```

## (2) INFORMATION FOR SEQ ID NO:179:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 460 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:179:

```

Met Thr Gln Ser Gln Thr Val Thr Val Asp Gln Glu Glu Ile Leu Asn
1          5          10          15

Arg Ala Asn Glu Val Glu Ala Pro Met His Asp Phe Phe Thr Asp Val
20         25         30

Pro Ile Thr Pro Cys Glu Leu Thr Ala Ala Lys Asn Ala Ala Gln Gln

```

[96]

35	40	45
Leu Val Leu Ser Ala Asp Asn Met Arg Glu Tyr Leu Ala Ala Gly Ala		
50	55	60
Lys Glu Arg Gln Arg Leu Ala Thr Ser Leu Arg Asn Ala Ala Lys Ala		
65	70	75
Tyr Gly Glu Val Asp Glu Gln Ala Ala Thr Ala Leu Asp Asn Asp Gly		
85	90	95
Glu Gly Thr Val Gln Ala Glu Ser Ala Gly Ala Val Gly Gly Asp Ser		
100	105	110
Ser Ala Glu Leu Thr Asp Thr Pro Arg Val Ala Thr Ala Gly Glu Pro		
115	120	125
Asn Phe Met Asp Leu Lys Glu Ala Ala Arg Lys Leu Glu Thr Gly Asp		
130	135	140
Gln Gly Ala Ser Leu Ala Ala Phe Ala Asp Gly Trp Asn Thr Phe Asn		
145	150	155
Leu Thr Leu Gln Gly Asp Val Lys Arg Phe Arg Gly Phe Asp Asn Trp		
160	165	170
Glu Gly Asp Ala Ala Thr Ala Cys Glu Ala Ser Leu Asp Gln Glu Arg		
180	185	190
Glu Trp Ile Leu His Met Ala Lys Leu Ser Ala Ala Met Ala Lys Gln		
195	200	205
Ala Glu Tyr Val Ala Gln Leu His Val Trp Ala Arg Arg Glu His Pro		
210	215	220
Thr Tyr Glu Asp Ile Val Gly Leu Glu Arg Leu Tyr Ala Glu Asn Pro		
225	230	235
Ser Ala Arg Asp Gln Ile Leu Pro Val Tyr Ala Glu Tyr Glu Gln Arg		
240	245	250
Ser Glu Lys Val Leu Thr Glu Tyr Asn Asn Lys Ala Ala Leu Glu Pro		
255	260	265
Val Asn Pro Pro Lys Pro Pro Pro Ala Ile Lys Ile Asp Pro Pro Pro		
270	275	280
Pro Pro Glu Glu Gln Gly Leu Ile Pro Gly Phe Leu Met Pro Pro Ser		
285	290	295
Asp Gly Ser Gly Val Thr Pro Gly Thr Gly Met Pro Ala Ala Pro Met		
300	305	310
Val Pro Pro Thr Gly Ser Pro Gly Gly Gly Leu Pro Ala Asp Thr Ala		
315	320	325

```

Ala Gln Leu Thr Ser Ala Gly Arg Glu Ala Ala Ala Leu Ser Gly Asp
340                               345                               350

Val Ala Val Lys Ala Ala Ser Leu Gly Gly Gly Gly Gly Gly Val
355                               360                               365

Pro Ser Ala Pro Leu Gly Ser Ala Ile Gly Gly Ala Glu Ser Val Arg
370                               375                               380

Pro Ala Gly Ala Gly Asp Ile Ala Gly Leu Gly Gln Gly Arg Ala Gly
385                               390                               395                               400

Gly Gly Ala Ala Leu Gly Gly Gly Gly Met Gly Met Pro Met Gly Ala
405                               410                               415

Ala His Gln Gly Gln Gly Gly Ala Lys Ser Lys Gly Ser Gln Gln Glu
420                               425                               430

Asp Glu Ala Leu Tyr Thr Glu Asp Arg Ala Tyr Thr Glu Ala Val Ile
435                               440                               445

Gly Asn Arg Arg Arg Glu Asp Ser Lys Glu Ser Lys
450                               455                               460

```

## (2) INFORMATION FOR SEQ ID NO:180:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 277 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:180:

```

Ala Gly Asn Val Thr Ser Ala Ser Gly Pro His Arg Phe Gly Ala Pro
1           5           10           15

Asp Arg Gly Ser Gln Arg Arg Asn Arg His Pro Ala Ala Ser Thr Ala
20           25           30

Thr Glu Arg Cys Arg Phe Asp Arg His Val Ala Arg Gln Arg Cys Gly
35           40           45

Phe Pro Pro Ser Arg Arg Gln Leu Arg Asn Arg Val Ser Arg Glu Ala
50           55           60

Thr Thr Arg Arg Ser Gly Arg Arg Asn His Arg Cys Gly Tyr His Pro
65           70           75           80

Gly Thr Gly Ser His Thr Gly Ala Val Arg Arg Arg His Gln Glu Ala

```



88										96										95														
Arg	Asp	Gln	Ser	Leu	Leu	Leu	Arg	Arg	Arg	Gly	Arg	Val	Asp	Leu	Asp																			
160															165										116									
Gly	Gly	Gly	Arg	Leu	Arg	Arg	Val	Tyr	Arg	Phe	Gln	Gly	Cys	Leu	Val																			
117															120										125									
Val	Val	Phe	Gly	Gln	His	Leu	Leu	Arg	Pro	Leu	Leu	Ile	Leu	Arg	Val																			
130															135										140									
His	Arg	Glu	Asn	Leu	Val	Ala	Gly	Arg	Arg	Val	Phe	Arg	Val	Lys	Pro																			
145															150										155									
Phe	Gln	Pro	Asp	Tyr	Val	Phe	Ile	Ser	Arg	Met	Phe	Pro	Pro	Ser	Pro																			
165															170										175									
His	Val	Gln	Leu	Arg	Asp	Ile	Leu	Ser	Leu	Leu	Gly	His	Arg	Ser	Ala																			
180															185										190									
Gln	Phe	Gly	His	Val	Glu	Tyr	Pro	Leu	Pro	Leu	Leu	Ile	Glu	Arg	Ser																			
195															200										205									
Leu	Ala	Ser	Gly	Ser	Arg	Ile	Ala	Phe	Pro	Val	Val	Lys	Pro	Pro	Glu																			
210															215										220									
Pro	Leu	Asp	Val	Ala	Leu	Gln	Arg	Glu	Val	Glu	Ser	Val	Pro	Pro	Ile																			
225															230										235									
Arg	Lys	Val	Arg	Glu	Arg	Cys	Ala	Leu	Val	Ala	Arg	Phe	Glu	Leu	Pro																			
245															250										255									
Cys	Arg	Phe	Phe	Gln	Ile	His	Gln	Val	Gly	Phe	Thr	Gly	Arg	Gly	His																			
260															265										270									
Pro	Arg	Arg	Ile	Gly																														
275																																		

## (2) INFORMATION FOR SEO ID NO:191:

1.2. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 192 amino acids  
(B) TYPE: amino acid  
(C) STRANDNESS:  
(D) TOPOLOGY: linear

1411 SOURCE: DESCRIPTION: SRC ID NO: 1411

Arg Val Ala Ala Ser Phe Ile Asp Trp Leu Asp Ser Pro Asp Ser Pro  
1 5 10 15

```

Leu Asp Pro Ser Leu Val Ser Ser Leu Leu Asn Ala Val Ser Cys Gly
      20                      25                      30

Ala Glu Ser Ser Ala Ser Ser Ser Ala Arg Ser Gly Asn Gly Ser Arg
      35                      40                      45

Tyr Thr Ser Met Pro Ser Gly Thr Arg Pro Gly Pro Arg Arg Ala Thr
      50                      55                      60

Ser Arg Asp Asp Arg Arg Ser Ala Thr Ser Val Ile Pro Ser Arg Arg
      65                      70                      75                      80

Ser Val Ala Pro Arg Ala Glu Phe Gly Thr Arg Leu Ala Ser His Arg
      85                      90                      95

Ala Ser Pro Ser Asn Ala Cys Pro Val Arg Ile Val Thr Ser Ala Ser
     100                      105                      110

Gly Arg Pro Ile Ser Ser Pro Pro Ile Val Arg Ser Arg Ser Cys Val
     115                      120                      125

Asp Lys Asn Gly Arg Arg Cys Ala Ser Gly Tyr Arg Arg Leu Asn Arg
     130                      135                      140

Ala Arg Ser Ser Ser Ile Ala Ala Arg Cys Arg Thr Ile Gly Thr Phe
     145                      150                      155                      160

Arg Arg Ser Arg Tyr Ser Ala Ser Met Asn Val Ser Thr Asn Ser Pro
     165                      170                      175

His Val Thr His Gly Val Ala Pro Gly Val Thr Arg Arg Ile Gly Gly
     180                      185                      190

```

## (12) INFORMATION FOR SEQ ID NO:182:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 196 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:182:

```

Gln Glu Arg Pro Gln Met Cys Gln Arg Val Ser Glu Ile Glu Pro Arg
 1              5              10              15

Thr Glu Phe Phe Asn Arg Cys Ala Leu Pro His Tyr Trp His Phe Pro
20              25              30

Ala Val Ala Val Thr Ser Lys His Ala Ser Leu Arg Glu Leu Ala Pro

```

35	40	45
Arg Asn Pro Arg Arg Ser Ser Arg Arg Asp His Glu Asp Arg Arg Val		
50	55	60
Ile Phe Ala Ala Thr Leu Val Ala Val Asp Pro Pro Leu Arg Gly Ala		
65	70	75
Gly Gly Glu Ala Asp Gln Leu Ile Asp Leu Gly Val Cys Arg Arg Gln		
85	90	95
Ala Gly Arg Val Arg Arg Gly Gln Glu Leu His His Arg His Arg His		
100	105	110
Gln Gly Ala Ala Pro Asp Leu Arg Arg Arg Arg Arg His Arg Arg Val		
115	120	125
Gln Gln His Arg Arg Leu Gln Arg Val Arg Gln Leu Arg Arg Tyr Val		
130	135	140
Glu Thr Ala His His Arg Arg Phe Ala Arg Thr Asp Arg Val Arg His		
145	150	155
His Val Arg Gly Phe Ser Asn His Arg Arg Arg Val Tyr Arg Gly		
160	170	175
Arg His Ser Gly Ala Gly Gly Cys Pro Ala Gly Gly Ala Gly Ser Val		
180	185	190
Gly Gly Ser Ala		
195		

## (2) INFORMATION FOR SEQ ID NO:153:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 211 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

## (a) SEQUENCE DESCRIPTION: SEQ ID NO:153:

Val Arg Cys Gly Thr Leu Val Pro Val Phe Met Val Gln Phe Leu Thr		
1	5	10
Ser Thr Asn Ala Pro Ser Leu Pro Ser Ala Tyr Ala Glu Val Asp Lys		
20	25	30
Leu Ile Gly Leu Phe Ala Gly Thr Ala Lys Arg Trp Ile Asn Gly Tyr		
35	40	45

Glu Arg Gly Gly Lys Asp His Pro Pro Ile Leu Arg Val Thr Pro Gly  
 58 55 60  
 Ala Thr Pro Trp Val Thr Trp Gly Gln Phe Val Glu Thr Arg Met Leu  
 65 70 75 80  
 Ala Glu Tyr Arg Asp Arg Arg Lys Val Pro Ile Val Arg Gln Arg Ala  
 85 90 95  
 Ala Ile Glu Glu Leu Arg Ala Arg Phe Asn Leu Arg Tyr Pro Leu Ala  
 100 105 110  
 His Leu Arg Pro Phe Leu Ser Thr His Glu Arg Asp Leu Thr Met Gly  
 115 120 125  
 Gly Glu Glu Ile Gly Leu Pro Asp Ala Glu Val Thr Ile Arg Thr Gly  
 130 135 140  
 Gln Ala Leu Leu Gly Asp Ala Arg Trp Leu Ala Ser Leu Val Pro Asn  
 145 150 155 160  
 Ser Ala Arg Gly Ala Thr Leu Arg Arg Leu Gly Ile Thr Asp Val Ala  
 165 170 175  
 Arg Leu Arg Ser Ser Arg Glu Val Ala Arg Arg Gly Pro Gly Arg Val  
 180 185 190  
 Pro Asp Gly Ile Asp Val His Leu Leu Pro Phe Pro Asp Leu Ala Asp  
 195 200 205  
 Asp Asp Ala Asp Asp Ser Ala Pro His Glu Thr Ala Phe Lys Arg Leu  
 210 215 220  
 Leu Thr Asn Asp Gly Ser Asn Gly Glu Ser Gly Glu Ser Ser Glu Ser  
 225 230 235 240  
 Ile Asn Asp Ala Ala Thr Arg Tyr Met Thr Asp Glu Tyr Arg Glu Phe  
 245 250 255  
 Pro Thr Arg Asn Gly Ala Gln Arg Ala Leu His Arg Val Val Thr Leu  
 260 265 270  
 Leu Ala Ala Gly Arg Pro Val Leu Thr His Cys Phe Ala Gly Lys Asp  
 275 280 285  
 Arg Thr Gly Phe Val Val Ala Leu Val Leu Glu Ala Val Gly Leu Asp  
 290 295 300  
 Arg Asp Val Ile Val Ala Arg  
 305 310

(2) INFORMATION FOR SEQ ID NO:154:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3072 base pairs

(B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xii) SEQUENCE DESCRIPTION: SEQ ID NO:184:

CTCTGTGCGA TTGCGCAGGA GCTGAGGAGC CCAMGGGGCC GTTCGGGAA GTCTGCGAGG	60
CATTGCGCGA CGGGCTGGCC GGCAGGGTA AGCAATCAA CAGCAGCTCG AACAGCCTCT	120
CGCAGGCGTT GAACCTCTG AATGAGGCGC GGGGGACTT CTTCGCGTC GTACCGAGCC	180
TAGCGCTATT GGTCAACGCG CTALATCAGG AGGACCAACA GTTCGTGCGC TTGACCAAGA	240
ACCTTGGCGA GTTCAAGGAC AGCTTACGCC ACTCGATTC GAACTCTGCG AACGCCATCC	300
AGCAATTCGA CAGCTTCTCG GCGTCCGCG GCGGCTCTT GCGCAAGAC CCGAGCTTC	360
TGACCGATGA CTTCAATAT CTGCGAGCGG TGGCAGCCAC GTTGTGAGC CGCGATCGGT	420
TGGATGGGTT GAGAGAGGTC CTGCAATCT TCGGAGGCT GCGGGCGAG ATTAAAGGCG	480
TTTAAGCTCC GACACAGGAT GCGTGGGCT GCGTTTCGC GTTCAAGAA TTGCGCAAC	540
CGATGGAGTT CACTTGCAGC TCGATTCAGG CGGTAGAGC GCTCGATTAT CAGAGCTGCG	600
CGGACCTCG TCGCGAGTAT CTGGCGCCAG TCGTGTATCG CATCAAGTC AACCTACTTC	660
CGTTGGGCTT GAGCTGCGC AGGCGCGCT CGACACTGCC TAAAGAGTC GGTACTTCC	720
AGCGCGGCTT GAGCGCGCC AACGGTACA AGGACACCG GGTGCGCGCC ATCTGGGTGC	780
CGGATACGCG GTTATCACAC CGCAACACCG AGCTGGGTC GTTGGTGGC CCGGCGTGC	840
AAGGGTTCA GGTGGAGCG ATCAGCGAG GTTCTCTAC GCGGAGTCC GTGGCGGAC	900
TCATGGGTGG TCGGATATG GCGCTCGCT GTCAGGCGT GCAAGCGCG CCGGAGCGCC	960
CGAATGGGTA CGACGAGTAC GCGCTGCTCG CGGATCGCG GTTACAGGCC CCGAGGTC	1020
CGAATACACC GCGGCTCTT GGGCGGAGC TAAAGCGCG TGGGTCGTA CAGGTCTTGG	1080
CGGGATCTGT GTTCCGAGA GATGCGCGCG CAGCTCGGAG AAGCTTCGAC TACATGGGCG	1140
TCTTGTGCT GTCGCGGCG CTGGGAGCT TCGTCTTGG GGTGTATCT AGCGGCGCG	1200
GTGAGGATAT GCGGATCGC CAGCTGTGA TACCGGATAT CAGCGGCTG GCTTCTCTCG	1260
CGGATCTGT CCGCATTCG TGGTACCGG CAGAGATTC GCTGAGAGC ATGCGCTTGT	1320
TCCAGAGGAG AGCGGTGCG CAGGCTAGA TGAGGATGAC AGTGCCTCT CTGCGGCTGT	1380

```

TTGGCTGCTT CTTCCTGCTC CCGAGCTACC TCGAGCAGST GTTCGACCAA TCACCGATGC 1440
AATCGGGGGT GCATATGATC CCACAGGGCC TCGTGGCAT GCTGGCGATG CCGATCGGCC 1500
GAGCGATGAT GGACCGAGCG GGACCGGCCA AGATCGTGCT GTTGGGGATC ATGCTGATCG 1560
CTCGGGGGTT GGGACCTTC GCGTTTGGTG TCGCGGGCA ACCGAGTAC TTACCGATTC 1620
TCCGACCGG GCTGGCAATC ATGGGATCG GCATGGGCTG CTTCATGATG GAGCTGTCCG 1680
GGGGGGCAAT GGAGACCGTC GCGCCAGATC AGATCGCTCG GAGTTCGAGC CTGATCAGCG 1740
TCAACCGACA GGTGGCGGT TCGATAGGA CCGACTGAT GTGCTGCTG CTCACCTAGC 1800
AGTTCAATCA CAGCGAAATC ATCGCTACTC CAAAGAAAGT CCGACTGACC CCGAGAGTGC 1860
GGCGGGGGCG GGGGGGGCG GTTGACCTT CCGCGCTACC GCGCGAATC AACTTGGCG 1920
CGCAACTGCT GCAATACCTT TCGACCGCT ACCGGGTGT ATCGTGATA GCGACCGGCG 1980
TAGTGGTCTC GAGCTGATC CCGCGGCAT TCGTGGGA ACAGCAGGT AGTCATCGAA 2040
GAGCACCGTT GCTATCGCA TGAAGTCTC TT
2092

```

(2) INFORMATION FOR SEQ ID NO:185:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1923 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:185:

```

TCACCTCGGA GAAGTCGCTC CTCGACGACC TGGACATGGA CTCGTGTGCG ATGCTCGAGA 60
TGGCGTGCA CACCGAGGAC AGTACGGCG TCAGGATCG CAGCGAGGAC CTCGCGGCTC 120
TGGGATCACT CCGTGAGTTT CTCGCTTACA TCGGAGGCT CGAGGAGGA AACCGGAGG 180
CGGTCAGGCG GTTGGCGCGG AGATTGAGT CGGAUAACCT CGATCGCGCA CAGCGAGATC 240
GTTGGGTTTC AGCGAGTCC CAGCTGCGG AGCGCGTCC TCGCTTTGCA CGCTCAGCCA 300
GTTTGGCGTG TCGGGCGTT CCGGAAATG TTCCGACCG AGGAGGGAC CCGTGGGAA 360
GTTGACTGAT CCGCGAGCA CATGCTGAT CCGACGCTG CTGACGATTC CCGCGGCTCC 420
GAGTTGGCGG GGGCGGATT GCGGCTTCC GTGAGAGCC AGCGGATCCC GCGGCGCGCC 480

```

GGCCGTGGCT	GGTCTTTTGG	GGCCGCCGAT	GGCCAGGACG	AGAACGACGA	TGGCGGGGAT	540
GAAGAGGCTC	AGCGCAATGA	CGACCGAGAG	ATTTCGCAGG	CATACCTCTT	GGTACCGGCTG	600
CGCCGGGGTT	GGTCGATCGG	TGCGATATCG	ATGGCGGCGT	TTAAGGTAAC	AGCTTTGGCG	660
GGAGCGGGGG	TGACAAAGGG	CGATTGTCTG	GGCGGGGAGC	CGGCGAGGTC	TGGCGCGCGG	720
TGACGTCAGC	TGACTGGGTC	ACGATCGCGG	TCTCGGTGAG	CGTGCACCTG	AAACACAGCTC	780
AAGGCGAAGT	GTTCCTCAGG	TGACGAGGTC	AAGCTCGACC	CGGATCTGCT	GGTACGTTTC	840
GAGCGCGGCG	AGGTGCGGAG	TGACGAGGTC	TGCGCGGACA	GGTTTCGCGG	TGAAGCGGAC	900
CAGGCGATCG	TGGTTTGGCG	CGCGGTGAGC	ATCGTCTCTG	CGGAGGTTGT	CGGTCAAGGC	960
GGGATATGAG	CAGGCGATCG	GTGCGAGGTA	GTTCCTGGAG	GTGATGTGCG	CGAAGTAGGC	1020
GTGACGCGCA	AGAGGGGCAA	TGCGATGGGG	CGCTGTATGC	GGGCTCAGGA	CGGAAACAGT	1080
TTGACGAGCC	GGTGGCGGCA	TGAGATGGAG	GGCAGCGTTG	AGCGCGCGCA	CGGCGGCGTC	1140
GTGCGCTTGG	TGCGAGGTCG	CGAATCGGCG	AAGCAGGCGG	CTGCTGTCTG	GTGCGATCAC	1200
CGCGGTGTGC	GATCGAGGCT	TTCGCGAGAG	ATTGCTTGGG	TGACCGGGGG	CAGGCGAGCT	1260
TGTGCGCTTG	CGGCGAGAAC	CGAGCTTCTG	CGAAGCGGTT	CGACACCGGT	CGGCGCGGCG	1320
TCAATCTGGA	TGCGCGGATG	GGGCTGGGTC	ATCTGCACTT	GGTGGTTTCC	CGCGAGGCTA	1380
AGCGCGCTCG	GAATCGCGCT	GGGATGACCC	AGAGCTGCTG	CGACATGCTT	GGTTGTTTCC	1440
ATGGTAGGAA	ATTGACATC	CGAGCTTCTG	TAGCGCTGTC	CTGCTCGGGA	TGTCGGGAGG	1500
ATCGCTGAGC	GTAGCGAGAG	ATTGTTTGGG	AAATGGCTGA	GGGAGCGTGC	GGTGGGGGTC	1560
ATCGCTGTGC	ATCGCGGCTT	AGCGCGATGC	GGGCTATCGG	TGATGAGAGC	TGGCGGTGGT	1620
CGGCGAGCTA	CGGCGCTGGA	TGTCGAGGTC	GTGCGCAGAC	CGTGGGATGC	GGCTTTGCGG	1680
CGGCGGCTGT	TGGCGATGAG	CGGCTGCTGC	GAGCACTGCG	TGGACAGGTA	TGATCGGAGG	1740
GTGCTGGCTA	TGCGAGGCTT	GTCTGTGAGC	CTGACCTGTA	CGAGGCTGAT	GGGCGCGCGG	1800
CAGGCGCGCG	CGGCTGAGCG	GGTGGCGGCG	CGGAGAGGTC	GTGTCGAGCT	GGATTTCGAT	1860
AGCGCGCGCG	AGGTCAGGCG	GGGCTGAGCT	GGGAGAGGTT	CGGCGAGGTA	GGCTCAGGTC	1920
ACC						1923

(2) INFORMATION FOR SEQ ID NO:186:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1855 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO:186:

CTGGGGTGGC AGTGTACAGC GCGATATGAC GTGGGCTTC ATTTCGGCG	60
CCCCGCCCCA CCAATCTG ACCACCGAGT CCGTCAGTTC CGAAGGTCC	120
GCCCAATATG GTGCTGGGTT TCTGGAACG AGCCCTGCCG TATCGGTGC	180
AACAACGACA GTCCAGGAAT CCGTCCCGC GCGGCGGATT CCGAGGGCA	240
GCGGATGTTT CAGAGGTAC GTTGGCTAC GTGAAAGCC GCGCCAGGC	300
GTCATGCA GCGAATATG AAAACCGCG ATATTGTCC GAGATTGTC	360
AGTCTTCT GTTCTGTTT TGGCAATGA GTTGCATGT GTTGAAGCC	420
GGGACAAAG TCTTTTGGC TTTCGTCGC CTTTACCGC CCGACGGGT	480
CGGCTGACT CCGCAATCT CCGCGGTTT TCGTTGCGC TCAGCTGGG	540
CCACCGGATG ATTATGACA CTGCGCGCT GCGCGGAGG ACGCGGCGA	600
CAGCGGCGG AAGGAGCGA CCGAGAGGC GCGGCGATG ACGAGTGA	660
GCTGGAACG AGTGGGTGG GAGAGCGCT GAACCGGCT TTGAGTGCC	720
AGCAGGTGA TTCCGCAATC TCGGCGGCG TGCTAGGAG GCGGCGGAG	780
TTTGACGAT GATCGCGGT CTGGGTGTT CCGCGGCGG GTATGACAG	840
CATGACAAAT TCGAGTAT AGGTCCAGT TCAGCAAGG CACAGGCAAC	900
GTTTATGAC GATCCGAC GCGATCGGG ACATGCGGG GCTTTTGA	960
AGACGTTGA GACAGAGCT CCGCGGATG GCGGTGCGG GCAAAACATC	1020
GCTGAGTGG CATGAGGAG CCGAGCTGCG TAGA	1055

## (2) INFORMATION FOR SEQ ID NO:187:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 259 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear



## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:187:

CGGCGTCGTT GTTGGCATAC TCTCCCGGGG GCGGCTCGAC GCGCTGGGCC GTGGCGTGTG	60
TCCGCGTGA CACCGCGGAT CCGCGAACCA TCCGAGATCA CCTCGCAATG ATCCACTTCG	120
CUCAGCTGAT CACCGAGCCA CGGCGGGTGT TCGGAGAGAG CTTGCGTCAAC CTTCGTATAG	180
CGTCCGGCCC CGAGCCGACG GAAGTTGTAG TACTGGCCCC CACCGCTGTT ACCGCGACCG	240
GAGAAATCA GGGTCAGGAT CGCATGTGG CGGCGGAGCT AGTCACCCCG GAAACACAGA	300
TCTTCGAGCA GGTGCTCGCG CGGCGGCCAC AGACCAAAAC CACCGCGGCG ATAGGTGAG	360

## (2) INFORMATION FOR SEQ ID NO:188:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 350 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:189:

ANCGGGCCCG TGGGCANNNG TCTCTAAGG GCTCTGTTG GTGGCATGAA GTGTGGGAAG	60
GATGATCTTT GCGAGATTCC GCGGAGAGCA AAGAGAGTCC TAGTCTTAGT CCGAGTCCGC	120
CGCAAGTCTC CTGGAATAC TCGGTACCG GAGGCGCRAA CCGGGTCTTC TTGCTAAGC	180
TACCGAACC ACTTGAGGTT CCGGAGCTCG TTGAGCTCCA GAGCGATTCC TTGAGATGCG	240
TGATGATTC GCGGCGCTAG CGGGAATCG CTGCGGAGCG AGGTGATGTC AAGCAATGG	300
GTGGCGCTGA AGAGGTGCTC TAGGAGCTGT CTGCGAICGA GCACTCTCC	360

## (2) INFORMATION FOR SEQ ID NO:189:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 378 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:190:

Glu Gln Pro Lys Gly Pro Phe Gly Gln Val Ile Gln Ala Phe Ala Asp  
 1 5 10 15  
 Gly Leu Ala Gly Lys Gly Lys Gln Ile Asn Thr Thr Leu Asn Ser Leu  
 20 25 30  
 Ser Gln Ala Leu Asn Ala Leu Asn Gln Gly Arg Gly Asp Phe Phe Ala  
 35 40 45  
 Val Val Arg Ser Leu Ala Leu Phe Val Asn Ala Leu His Gln Asp Asp  
 50 55 60  
 Gln Gln Phe Val Ala Leu Asn Lys Asn Leu Ala Gln Phe Thr Asp Arg  
 65 70 75 80  
 Leu Thr His Ser Asp Ala Asp Leu Ser Asn Ala Ile Gln Gln Phe Asp  
 85 90 95  
 Ser Leu Leu Ala Val Ala Arg Pro Phe Phe Ala Lys Asn Arg Gln Val  
 100 105 110  
 Leu Thr Ala Asp Val Asn Asn Leu Ala Thr Val Thr Thr Thr Leu Leu  
 115 120 125  
 Gln Pro Asp Pro Leu Asp Gly Leu Gln Thr Val Leu His Ile Phe Pro  
 130 135 140  
 Thr Leu Ala Ala Asn Ile Asn Gln Leu Tyr His Pro Thr His Gly Gly  
 145 150 155 160  
 Val Val Ser Leu Ser Ala Phe Thr Asn Phe Ala Asn Pro Met Gln Phe  
 165 170 175  
 Ile Cys Ser Ser Ile Gln Ala Gly Ser Arg Leu Gly Tyr Gln Gln Ser  
 180 185 190  
 Ala Gln Leu Cys Ala Gln Tyr Leu Ala Pro Val Leu Asp Ala Ile Lys  
 195 200 205  
 Phe Asn Tyr Phe Pro Phe Gly Leu Asn Val Ala Ser Thr Ala Ser Thr  
 210 215 220  
 Leu Pro Lys Gln Ile Ala Tyr Ser Gln Pro Arg Leu Gln Pro Pro Asn  
 225 230 235 240  
 Gly Tyr Lys Asp Thr Thr Val Pro Gly Ile Thr Val Pro Asp Thr Pro  
 245 250 255  
 Leu Ser His Arg Asn Thr Gln Pro Gly Trp Val Val Ala Pro Gly Met  
 260 265 270  
 Gln Gly Val Gln Val Gly Pro Ile Thr Gln Gly Leu Leu Thr Pro Gln  
 275 280 285

Ser Leu Ala Glu Leu Met Gly Gly Pro Asp Ile Ala Pro Pro Ser Ser  
290 295 300

Gly Leu Gln Thr Pro Pro Gly Pro Pro Asn Ala Tyr Asp Gln Tyr Pro  
305 310 315 320

Val Leu Pro Pro Ile Gly Leu Gln Ala Pro Gln Val Pro Ile Pro Pro  
325 330 335

Pro Pro Pro Gly Pro Asp Val Ile Pro Gly Pro Val Pro Pro Val Leu  
340 345 350

Ala Ala Ile Val Phe Pro Arg Asp Arg Pro Ala Ala Ser Glu Asn Phe  
355 360 365

Asp Tyr Met Gly Leu Leu Leu Leu Ser Pro Gly Leu Ala Thr Phe Leu  
370 375 380

Phe Gly Val Ser Ser Ser Pro Ala Arg Gly Thr Met Ala Asp Arg His  
385 390 395 400

Val Leu Ile Pro Ala Ile Thr Gly Leu Ala Leu Ile Ala Ala Phe Val  
405 410 415

Ala His Ser Tyr Tyr Arg Thr Gln His Pro Leu Ile Asp Met Arg Leu  
420 425 430

Phe Gln Asn Arg Ala Val Ala Gln Ala Asn Met Thr Met Thr Val Leu  
435 440 445

Ser Leu Gly Leu Phe Gly Ser Phe Leu Leu Leu Pro Ser Tyr Leu Gln  
450 455 460

Gln Val Leu His Gln Ser Pro Met Gln Ser Gly Val His Ile Ile Pro  
465 470 475 480

Gln Gly Leu Gly Ala Met Leu Ala Met Pro Ile Ala Gly Ala Met Met  
485 490 495

Asp Arg Arg Gly Pro Ala Lys Ile Val Leu Val Gly Ile Met Leu Ile  
500 505 510

Ala Ala Gly Leu Gly Thr Phe Ala Phe Gly Val Ala Arg Gln Ala Asp  
515 520 525

Tyr Leu Pro Ile Leu Pro Thr Gly Leu Ala Ile Met Gly Met Gly Met  
530 535 540

Gly Cys Ser Met Met Pro Leu Ser Gly Ala Ala Val Gln Thr Leu Ala  
545 550 555 560

Pro His Gln Ile Ala Arg Gly Ser Thr Leu Ile Ser Val Asn Gln Gln  
565 570 575

Val Gly Gly Ser Ile Gly Thr Ala Leu Met Ser Val Leu Leu Thr Tyr

580	595	590
Gln Phe Asn His Ser Glu Ile Ile Ala Thr Ala Lys Lys Val Ala Leu 595	600	605
Thr Pro Glu Ser Gly Ala Gly Arg Gly Ala Ala Val Asp Pro Ser Ser 610	615	620
Leu Pro Arg Glu Thr Asn Phe Ala Ala Gln Leu Leu His Asp Leu Ser 625	630	635
His Ala Tyr Ala Val Val Phe Val Ile Ala Thr Ala Leu Val Val Ser 645	650	655
Thr Leu Ile Pro Ala Ala Phe Leu Pro Lys Gln Glu Ala Ser His Arg 660	665	670
Arg Ala Pro Leu Leu Ser Ala 675		

## (2) INFORMATION FOR SEQ ID NO:190:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 120 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:190:

Thr Pro Glu Lys Ser Phe Val Asp Asp Leu Asp Ile Asp Ser Leu Ser 1	5	10	15
Met Val Glu Ile Ala Val Gln Thr Glu Asp Lys Tyr Gly Val Lys Ile 20	25	30	
Pro Asp Glu Asp Leu Ala Gly Leu Arg Thr Val Gly Asp Val Val Ala 35	40	45	
Tyr Ile Gln Lys Leu Glu Glu Glu Asn Pro Glu Ala Ala Gln Ala Leu 50	55	60	
Arg Ala Lys Ile Glu Ser Glu Asn Pro Asp Ala Ala Arg Ala Asp Arg 65	70	75	80
Cys Val Ser Pro Thr Ser Gln Ala Arg Asp Ala Arg Arg Pro Leu Ala 85	90	95	
Arg Ser Ala Arg Leu Ala Cys Arg Arg Leu Pro Ala Ser Val Pro Thr 100	105	110	

Thr Arg Arg Asp Pro Arg Glu Arg  
115 120

## (2) INFORMATION FOR SEQ ID NO:191:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 55 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:191:

Leu Ala Cys Glu Cys His Arg Arg Tyr Asp Val Gly Ile Glu Phe Arg  
1 5 10 15  
Gly Pro Ala Gly Pro Val Ala Thr Gln Ser Gly Pro Pro Gly Pro Ser  
20 25 30  
Ile Ala Glu Gly Arg Glu Val Arg Ala Glu Cys Gly Ala Gly Phe Leu  
35 40 45  
Glu Arg Arg Pro Ala Val Ser Gly Ala Leu Pro Pro Asn Asn Ala Ser  
50 55 60  
Pro Gly Ile Arg Ser Arg Ala Ala Asp Ser Gln Arg His Leu Leu Ala  
65 70 75 80  
Gly Asp Gly Ser Asp Val Thr Val Gly  
85

## (2) INFORMATION FOR SEQ ID NO:192:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 119 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:192:

Ala Ser Leu Leu Ala Tyr Ser Ala Ala Ala Ala Ser Thr Ala Leu Ala  
1 5 10 15  
Val Ala Cys Val Arg Ala Asp His Arg Asp Arg Arg Thr Ile Arg Asp  
20 25 30

[illegible]

## 12. INFORMATION FOR SFO ID NO. 193:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 116 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:193:

Arg Ala Arg Gly His Arg Ser Ser Lys Gly Ser Arg Trp Ser His Glu  
1 5 10 15  
Val Leu Glu Gly Cys Ile Leu Ala Asp Ser Arg Glu Ser Lys Thr Ala  
20 25 30  
Ala Ser Pro Ser Pro Ser Arg Pro Glu Ser Ser Ser Asn Asn Ser Val  
35 40 45  
Pro Gly Ala Pro Asn Arg Val Ser Phe Ala Lys Leu Arg Glu Pro Leu  
50 55 60  
Glu Val Pro Gly Leu Leu Asp Val Glu Thr Asp Ser Phe Glu Trp Leu  
65 70 75 80  
Ile Gly Ser Pro Arg Trp Arg Glu Ser Ala Ala Glu Arg Gly Asp Val  
85 90 95  
Asn Pro Val Gly Gly Leu Glu Glu Val Leu Tyr Gln Leu Ser Pro Ile  
100 105 110  
Glu Asp Phe Ser  
115

## (2) INFORMATION FOR SEQ ID NO:194:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 911 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) SEQUENCE DESCRIPTION: SEQ ID NO:194:

```

TCTACGGCAG CAATGCGTTT GGTGACAGAT GTGATGCGG GCTTCGCTGC TGGGATGGC      60
GTGAAGAGCG CGACGCTGTT CGCGCGATTC GGGGAGAACA TCGACTGCTT CAAAGAGCTG      120
GTGCGGGGCG CAGTCGATCG GTTCGCGCAC GAGCGCACGT GCACGCGCTT TACACACCAC      180
GCGGATGTTC CGTTGCGGTT CGACGCGCGA TGAGGCTGCT GTTGACXGCG GCGGCGGCTT      240
TCATCGGCTC GCGCGTGATG GCGGCGTTAC GAGCTCGGCG TCAGGACGCG GTGGCGCTCG      300
AGCGCTGCTT GCGCGCGCGC CAGGCGGCAA ACTTGATGCT GCGGCGGCGT TCGGACGCGG      360
TGGAGCTCGG CGAGCGGAGC GCGCGGCGCG GTTGTGCGC GGTGTGCGAT CTGATGTGTC      420
ACGAGGCGCG CATGCTGAGT GCGGCGCTCA AGGCGCGCGA CGTACGCGCG TATGCGGCGC      480
ACAAGCATTT CAGCAGCAGC GTGCTGCTCG CGGATGCTT GCGCGCGGCG GTCGCGCGCTT      540
TGGTGCTGCG GTGCTGATG GTGCTTTAGG CGCAGCGGCG CTATGCTGCT GCGCAGCATG      600
GCGCGCTGCG CCGCTGCGCG GCGCGCGGCG CGGCTTGGG GAATGCGCTT TTGAGGAGCG      660
GTGCGCGGCG GTGCGCGGCG CAGTCATCTT GCGAATTGCT CGAGGAGAT GCGCGCTTGC      720
GCGCGCGGCG GGTGTACGCG GCGCAGGAGC GCGCGAGGAG CACTACGCGC TCGGCTGCTC      780
GGAGACGAGT GCGGCTGCGC TGTGCGGCTT G

```

## (2) INFORMATION FOR SEQ ID NO:195:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 965 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (01) SEQUENCE DESCRIPTION: SEQ ID NO: 195.

GCTCCGCVAT	ATGAGCCGAC	ATGACTTTAS	GCAACACAG	CGTAGTAGTC	GAATATATAC	40
GACTTTGTGG	TCCCGGTGGC	GAGATAGAGC	AGCTGTGCGC	GTTCGTGAGC	GTACCCGGTT	120
GCTGGAGAC	GGAACCCATG	CTTTCAACCT	AGCTGTGCGC	TCCACACAGT	CGGACAGGTA	180
AGCTGACGCT	CAATATACG	CTGCAATTTC	GCGGTGAGCT	TCCGCTGCGC	GACAATCAAG	240
CGTAGTCTAC	TTACTTCCGA	GCAATTTGAA	CGGTTTCGAT	CGGTTTCGCG	CTGGTGAAAG	300
TCCCGGTCAA	GCTGTACAGC	GCTACCGTAG	ACCACGACAT	CAGGTTTCAC	CAGGTGCACG	360
CCAGAGGCGA	CGAGCCGATT	CGGTACAAAG	GCGCTTCGGA	GCGGTGTGCG	CGGCTGTGCG	420
ACTACCGCGA	TCTTGGCCCG	GCTTACGAGT	CGGCGGACCG	CGAATGTGTC	GCGTTCACCG	480
AGGAGGACAT	GCGGAGCTTG	CTTGAGAGAC	GAGGCCCGGA	GTTCGAGGTC	TTGGAGTTCC	540
TCCCGCCGCG	CGAGCTGGAC	CGGATGATCT	TGAGCCGCGC	CTAGTTTTTG	GAGCTGATT	600
CGAGTCTCTC	GAATGCTAT	GTGCTGTGCG	CTAGGACACT	CGCCGAGAGC	GAGCGGATGG	660
CGATCTGCGA	TCCGCTCCAG	GCGGTGAGAT	CGAGGAGAAA	TGAGAGCGCG	TATCCACAGT	720
TCCGCTGCGA	GTCGCGCAC	CACAAACGCT	AGGAGGATCG	AGCATTTCCG	GAGCTGAAGT	780
GCGGCGCTAT	AGAGCGCGCT	CTCGCGGATT	ATCAAGACCG	AAATGCGCTT	ACTCATGCCA	840
TGAGCGCTCG	TGACCGGATG	GAGCTTTTTT	GCGGAGCTCC	AGCGCTGCGC	GCGCGACCTC	900
AAGTGGGCGT	GCAACCCGCT	CTTTCGCGGA	AAGCGGTTTG	GCGGCTGCGC	GTCATCTGCT	960
CTCTCT						96

## (2) INFORMATION FOR SSG ID NO: 196.

### 2.1 SEQUENCE CHARACTERISTICS:

- ```
(A) LENGTH: 2367 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
```

## 4832 UNKNOWN DESCRIPTION: SEQ ID NO: 198.

|             |           |           |           |            |            |     |
|-------------|-----------|-----------|-----------|------------|------------|-----|
| CCGCGGCGCCG | GCGAGTGGG | CAGCGCGCG | GTTTCCGCG | GTTCGGCGG  | TTGCGCGCG  | 46  |
| TTCCCGCGCG  | CGCGCGCGG | CGCGCGATG | ATTTTCATG | CGTAAAGTA  | CTGCGGTTTC | 126 |
| CAGCGGATG   | GCGGTTTGG | CGCGTATG  | CAGCGCGCG | GATTTCCGCG | GCGCGCGCG  | 186 |



|                                                                  |      |
|------------------------------------------------------------------|------|
| CTCCGCCGCT GGCACGTTG CCCCCTTGG CGATCAACAT GCGCTGCGC CCACCTTTG    | 240  |
| CACCGACCGC ACCGGCTCGG CCACTCCCGC GGACACCAAG CGAGCTGCGG CGGAGCGAC | 300  |
| CATCACCAAG TACGGACCGC ACCGCGGAGA CACCGCGAC CGGGCTTGG TGAACGTCG   | 360  |
| CGGTGCCAGC ACCGCGCGG TTACCGCGA CCGACCGCG NACCGCGCG CGCGCTCGG     | 420  |
| CGCGCGCGG CGCGCTTGG CGGTTCGCG GGTACCGGA CAACAACCG CGCGCGCGG      | 480  |
| CGTTCCCGG CGCGCGCGG GTCCGCGCG CGCGCGCGG GCGACCGCG CTGCGCGCT      | 540  |
| TGCGCGCGC ACCGCGTTG CCGCGCGCG CATCGCGTC TCGCGCGG TCTCGCGGT       | 600  |
| CAACGCTCG GATCGCGCG TTGCGCGCG TTGCGCGCG CCGCGCGCG CGCGCTTAC      | 660  |
| CACGCGCGC ACCGCGCGC CCGCGCGCG CGTTCGCGC ATCAGCGAT ACCGCGCGC      | 720  |
| CGCGCGCGC ATTGCGCGC GTTCGCGCG CGCGCGCGC CGCGCGCGC CGCGCGCGC      | 780  |
| CACGCGCGC ACCGCGCGC CCGCGCGCG CACGCGCGC GTTCGCGCG CGCGCGCGC      | 840  |
| CGCGCGCGC CGCGCGCGC CCGCGCGCG CACGCGCGC CGCGCGCGC CGCGCGCGC      | 900  |
| ACCGCGCGC ACCGCGCGC CCGCGCGCG CCGCGCGCG CGCGCGCGC CGCGCGCGC      | 960  |
| CGCGCGCGC CGCGCGCGC CCGCGCGCG CGCGCGCGC CGCGCGCGC CGCGCGCGC      | 1020 |
| CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC      | 1080 |
| CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC      | 1140 |
| CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC      | 1200 |
| CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC      | 1260 |
| CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC      | 1320 |
| CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC      | 1380 |
| CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC      | 1440 |
| CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC      | 1500 |
| CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC      | 1560 |
| CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC      | 1620 |
| CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC      | 1680 |
| CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC      | 1740 |
| CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC CGCGCGCGC      | 1800 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| CGGCCGACATT AGCCGCGCTC ACCTGCGAAG CAAATGCGAA TAAATCCAAA GCGGCAGTTG | 1890 |
| CCAGCAGCTG CCGGCTGGCG ATCAGCAGAG ACACCTGCGA CCTCCGGATA CCGCATATCG  | 1920 |
| CGGCACCGTG TCGCAGCGCG CACCGTGACC TTGAGTGGCF GCGTCGCGAG CCGGACTATG  | 1980 |
| GCGCGGACGG CGCTGCTCTT GATTGCGGCG GCGCGCGAGC TTGTTGGGCG AGTTGAGAC   | 2040 |
| GGGAGAGCAG GCGGAGCTTG CTGTAGAGCT GCGTCAGATG GGAATGACAG GTCCGCGGCG  | 2100 |
| AGATGAAATG (KAGAGCGG) ATGTCTTGT TGTGAGTCC CTCACCGAGC AGTAGAGCCA    | 2160 |
| CCTCAAGCTC TGTGCGCTTC AAGCGGCTTC AGCGCTTGT GCGGCGTTTC CGTGAGCGCG   | 2220 |
| GAGCTCGTTG CCGGTATGCG ATCGCTCAT CGATCGATAA CCGAGTCTCT TCGGCGCAGC   | 2280 |
| CTGTGCGAAT CTGCGTGTCA CCGATGCGTT TTGGAAGGCT GCGTAGAGAC GAGTTACAGC  | 2340 |
| CGGCTGGTA GATCGCGAGC CGGAGCG                                       | 2367 |

## (2) INFORMATION FOR SEQ ID NO:197:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 376 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:197:

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Gln | Pro | Ala | Gly | Ala | Thr | Ile | Ala | Ala | Ser | Ser | Pro | Cys | Ala | Thr | Val |
| 1   |     |     | 5   |     |     |     |     |     | 10  |     |     |     |     | 15  |     |
| Gly | Ala | Gly | Gly | Gly | Thr | Gly | Ser | Pro | Val | Thr | Thr | Glu | Thr | Ala | Ala |
|     |     | 20  |     |     |     |     | 25  |     |     |     |     |     |     | 30  |     |
| Thr | Thr | Gly | Arg | Gly | Gly | Ser | Gly | Asp | Val | Tyr | Gln | Ser | Ala | Ala | Ser |
|     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
| Gly | Ala | Ala | Ala | Thr | Thr | Pro | Thr | Ala | Gly | Gly | Tyr | Thr | Val | Gly | Pro |
|     |     | 50  |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| Val | Ala | Thr | Ile | Thr | Ala | Tyr | Gly | Ala | Arg | Asn | Val | Ala | Leu | Arg | Asp |
|     |     | 65  |     |     | 70  |     |     |     | 75  |     |     |     |     | 80  |     |
| Ser | Ala | Val | Ala | Ala | Val | Ala | Ala | Ala | Ala | Thr | Gly | Ser | Gly | Gly | Thr |
|     |     | 85  |     |     |     |     |     |     | 90  |     |     |     |     | 95  |     |
| Ala | Val | Thr | Thr | Gly | Thr | Ala | Gly | Gly | Leu | Ala | Ala | Ala | Cys | Arg | Arg |
|     |     | 100 |     |     |     |     | 105 |     |     |     |     |     | 110 |     |     |

Gly Gly Thr Val Ala Ala Gly Ala Thr Gly Arg Arg Ala Gly Ser Ala  
 115 120 125  
 Met Ala Ala Arg Ala Ala Val Ala Ala Gly Leu Ile Thr Asp Ala Gly  
 130 135 140  
 His Ile Cys Arg Ala Val Pro Gly Ala Gly Arg Gly Ala Gly Arg Gly  
 145 150 155 160  
 Ile Asp Pro Val Cys Pro Gly Glu Ala Gly Ala Ala Gly Thr Thr Gly  
 165 170 175  
 Ala Ala Met Ala Glu Gln Pro Gly Val Ala Ala Val Thr Ala Arg Thr  
 180 185 190  
 Pro Asp Ala Cys Gly His Ala Gly Ala Ala Asp Thr Ala Val Ala Ala  
 195 200 205  
 Val Ala Pro Gln Pro Pro Pro Val Pro Thr Gly Thr Ala Gly Arg Ala  
 210 215 220  
 Gly Thr Thr Gly Pro Ala Val Ala Ala Val Ala Asp Glu Pro Gly Arg  
 225 230 235 240  
 Ala Ser Ala Ala Ala Gly Leu Thr Glu Pro Ala Ser Arg Ala Val Ala  
 245 250 255  
 Thr Val Ala Lys Glu Glu Pro Ala Gly Arg Ala Arg Leu Pro Gly Cys  
 260 265 270  
 Arg Pro Val Gly Ala Val Ser Asp Gln Arg Ala Pro Gln Lys Arg Leu  
 275 280 285  
 Gly Gly Arg Ile His Arg Thr Gln Glu Thr Pro Leu Asn Ser Gly Pro  
 290 295 300  
 Ser Ala Gly Ile Pro Thr Arg Gly Arg Ser Gln Arg Leu His Lys Leu  
 305 310 315 320  
 Leu Val Lys Arg Cys His Leu Tyr Ala Glu Arg Leu Ile Leu Pro Ser  
 325 330 335  
 Met Gly Pro Gln Glu Pro Arg Asn Arg Arg Arg His Phe Ile Gly Ser  
 340 345 350  
 Arg Ser His His Phe Arg Arg Arg Asp Arg Arg Gly Arg Ile Ser Arg  
 355 360 365  
 Ala His Leu Arg Thr Asn Ser Arg  
 370 375

(2) INFORMATION FOR SEQ ID NO:198:

(1) SEQUENCE CHARACTERISTICS:

(a) LENGTH: 2852 base pairs

(B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:198:

|             |            |            |             |            |            |      |
|-------------|------------|------------|-------------|------------|------------|------|
| GGCCAAACAC  | CCCCCGCGAT | GGCGGCCACC | GGAGGCGCTT  | AGGACCAAGT | GTGAGGCCAG | 60   |
| GAGTGGCGCG  | CGATGTTTGG | CTACCATGCG | GGGGCTTCGG  | CGGCGCTCTC | GGCTTGACAA | 120  |
| CGGTTCGGCG  | AGGAGCTGCC | GACCGTACCG | GGGCGCGGTG  | CGGTGCTCAG | CGCGACCGCG | 180  |
| GTCCAGGTGA  | CGACCGGGCT | CTTCCGCAAC | CTGGGCTTGG  | CGAACGTCCG | CGAGGGCAAC | 240  |
| GTGGCGACCG  | GTATGTCTCG | GAATTCACAT | GTGGGCTTGG  | CGAACATCGG | CGACGGCAAC | 300  |
| ATCGGCGCGCG | GCAACATCGG | CAGCTCGAAC | ATCGGCTTTG  | GCAACGTGGG | TCTGTGGTTG | 360  |
| AGCGCAGCGC  | TGACACAGAT | CGTTTGGCG  | AAGCGCGCGC  | CGAACAGATG | CGGTTTGGCG | 420  |
| NACACCGCGC  | CGAACAGATG | CGGTTTGGCG | AATACGGGAG  | ACCGCAGCGC | AGTATCGGGG | 480  |
| CTCGCGCGCG  | CGGTTTGGTT | GGGTTTGGCG | GGCTTCAACT  | CGGCGACCGG | CGACATCGGT | 540  |
| CTGTTTCAACT | CGGCGACCGG | AAACGTGCGC | ATCGGCTCACT | CGGCGACCGG | CGACTTGGCG | 600  |
| ATGTCCAACT  | CGGCGACAGC | GTACACACCG | GGTTTGGCGA  | ACTCGCGCGA | CGGCAACACG | 660  |
| GGCTTCTTCA  | ATTCGCGAAT | AGGCAACACG | GGGCTCGCGA  | ACCGCGCGCA | GTACACACCG | 720  |
| GGTAGCTACA  | AGGCGCGCA  | CAGCAATACG | CGGCGCTTCA  | AGATGGCGCA | GTACACACCG | 780  |
| GGTCACTTGA  | AGGCGCGCA  | CTACACACCG | GGCTTGGCGA  | ACTCGCGCGA | TGTCACACCG | 840  |
| GGGCGCTTCA  | TGCTCGCGCA | CTTCACACCG | GGTTTCTTCT  | GGCGCGCGCA | CGACCAACGG | 900  |
| CGGATTTTCA  | GGGCGCGCGG | CTTCTTCAAC | TGACACAGTG  | CGGCGCTGTC | GGGATTTCTC | 960  |
| AGGACCGCTG  | CGGCGCGCGG | GTGCGGCTTC | CTAACCTTCC  | GTGCGACGAA | TTCGTGCTTC | 1020 |
| TTCACCTTCT  | CGGCGCGCGG | CATCGGTACG | TGCGGCTTGG  | CAACCGCGCG | GTGCTGCTTA | 1080 |
| TGCGGCTTGA  | TGACCTCGCG | CGACACCGTA | TGCGGTTTGT  | TGACATGAGG | CGTGGTGGCG | 1140 |
| ATGACACCGC  | CGGCTTGTAT | CTCGGCTTTC | TTCACACCGC  | GAACAGACAT | GTGCGGCTTT | 1200 |
| TTCGCTGCTC  | CACCGCTTCT | CAATCTCGCG | CTGCGACACG  | GGGCGCTGCT | GAACATTTCT | 1260 |
| GGACACCGCA  | AGATCGCGCA | TTCACACATT | CTCGGCGACG  | GAACGCTCGG | TGACTTCAAC | 1320 |
| ATGCTTGGCA  | CGGCGACGCT | GGGCGCGCA  | GAACATTTCT  | CGGCGCGCA  | CGTGGCGCGG | 1380 |

|            |            |            |            |           |            |      |
|------------|------------|------------|------------|-----------|------------|------|
| TTCAATATCG | GCAGTGGAAA | CATCGSAGTA | TTCAATGTGG | GTTCGGGAG | GCTGGAAAC  | 1440 |
| TACAACTCG  | GATCGGAAA  | CCTCGGATC  | TACAACTCG  | GTTTGGAAA | GTCGGGGAC  | 1500 |
| TACAACTCG  | GCTTGGGAA  | CGCGGGGAC  | TTCAACCAAG | GCTTGGCAA | CACGGGCAC  | 1560 |
| AACAACATCG | GCTTGGCAA  | CACCGCAAC  | AACAACATCG | GCATCGGCT | GTCGGGGAC  | 1620 |
| AACCGGAGG  | GCTTCATAT  | TGCTAGGGG  | TGCACTCGG  | GCATCGGAA | GAGGGGGCTG | 1680 |
| TTCAATTCG  | GCACCAATA  | GCTTGGGATC | TTCAACGGG  | GCACGGAAA | GCTCGGGATC | 1740 |
| GCAACTCGG  | GCACGGGAA  | CTGGGATATC | GGGAGCGG   | GTACCGCAA | TACGGGATC  | 1800 |
| CTCAATGCG  | GCACCTACAA | CGCGGGATC  | CTCAACGGG  | GCACCTCAA | CACGGGCTTC | 1860 |
| TACAACTCG  | GCACCTACAA | CACCGGGGG  | TTCAACGGG  | GTACCACAA | CACCGGCAC  | 1920 |
| TTCAACTCG  | GTCAACCAA  | TACCGGGG   | TATAACGGG  | GTCAACCAA | CACGGGCTTC | 1980 |
| TTCAATCGG  | GCACCTCAA  | TACCGGGG   | TTCAACGGG  | GCACCTCAA | CAATGGCTTC | 2040 |
| TTGGTGGGG  | GGGATACAA  | GGGCGAGAT  | GCATCGATC  | TCTGGGATC | CATTCGCTTC | 2100 |
| ATCGGCAAA  | ACGAGGAGT  | GTCAATGAG  | GTACCAAGG  | TATGAGCTT | GGGGGGCAAC | 2160 |
| ATGATCGG   | TCACGGGAG  | CTGAGGGGT  | TTCCGCGAA  | GCTTCATCT | GAGGGGTTG  | 2220 |
| TTCTGCTCG  | GGCGGGTAA  | TTCAACGGG  | TCCACGGGA  | GGTTCGAGG | GATCAAGCTC | 2280 |
| ACCATCGGG  | GACCGAGAT  | GACCTCGGG  | ATCAAGATG  | TGGGGGCTT | GGAGAGCGG  | 2340 |
| ACCATTCCT  | TCTCAAGAT  | CGATCGGGG  | CGGGGATCG  | GAAATCGAG | CACCAACCG  | 2400 |
| TCTCGGGCT  | TCTCAACTC  | GGGCGGGGT  | GGCAATCTG  | GCTTCGAAA | GCTGGGGGG  | 2460 |
| GCATCTTCA  | GCTCTGAAA  | CAGTGGTTC  | AGCAGCGGA  | TAGGGAATC | GCTTTTCAA  | 2520 |
| AACCTCGGT  | CGCTCGAGT  | AGCTGGGGG  | AGCTGGGGA  | ACTCGGATC | GCGCTTTTC  | 2580 |
| AACAGGATA  | CGCTGAAGT  | CTGAGGGGG  | GCAATGCTT  | CGGGGTTAA | CAACATCGG  | 2640 |
| AGCAACTCT  | CGCGGTTTC  | CGCGGGTGG  | ACCGGACGA  | TTTCAAGGC | GGGCTTTGG  | 2700 |
| AACCTGGGT  | AGCTGAAGT  | CGCGGGGGG  | TCTGGGGGA  | TTTGGAGGA | GTTAGATAG  | 2760 |
| GTTTCAACA  | TGATATCGG  | GTTTTCGGG  | AGTGCATCG  | AGCAATCGA | CGCGGAGGC  | 2820 |
| GTAGCGAAT  | AGCGGAGTC  | GCGGCTTTC  | AT         |           |            | 2880 |

(2) INFORMATION FOR SEQ ID NO:199:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 943 amino acids

(B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:199:

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Gly | Gln | Asn | Ala | Pro | Ala | Ile | Ala | Ala | Thr | Gln | Ala | Ala | Tyr | Asp | Gln |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
| Met | Trp | Ala | Gln | Asp | Val | Ala | Ala | Met | Phe | Gly | Tyr | His | Ala | Gly | Ala |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     |     | 30  |     |
| Ser | Ala | Ala | Val | Ser | Ala | Leu | Thr | Pro | Phe | Gly | Gln | Ala | Leu | Pro | Thr |
|     |     |     | 35  |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
| Val | Ala | Gly | Gly | Gly | Ala | Leu | Val | Ser | Ala | Ala | Ala | Ala | Gln | Val | Thr |
|     |     | 50  |     |     |     | 55  |     |     |     |     |     | 60  |     |     |     |
| Thr | Arg | Val | Phe | Arg | Asn | Leu | Gly | Leu | Ala | Asn | Val | Arg | Gln | Gly | Asn |
|     |     | 65  |     |     | 70  |     |     |     |     | 75  |     |     |     | 80  |     |
| Val | Arg | Asn | Gly | Asn | Val | Arg | Asn | Phe | Asn | Leu | Gly | Ser | Ala | Asn | Ile |
|     |     |     | 85  |     |     |     |     | 90  |     |     |     |     |     | 95  |     |
| Gly | Asn | Gly | Asn | Ile | Gly | Ser | Gly | Asn | Ile | Gly | Ser | Ser | Asn | Ile | Gly |
|     |     | 100 |     |     |     |     | 105 |     |     |     |     |     | 110 |     |     |
| Phe | Gly | Asn | Val | Gly | Pro | Gly | Leu | Thr | Ala | Ala | Leu | Asn | Asn | Ile | Gly |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| Phe | Gly | Asn | Thr | Gly | Ser | Asn | Asn | Ile | Gly | Phe | Gly | Asn | Thr | Gly | Ser |
|     |     | 130 |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| Asn | Asn | Ile | Gly | Phe | Gly | Asn | Thr | Gly | Asn | Gly | Asn | Arg | Gly | Ile | Gly |
|     |     | 145 |     |     | 150 |     |     |     | 155 |     |     |     | 160 |     |     |
| Leu | Thr | Gly | Ser | Gly | Leu | Leu | Gly | Phe | Gly | Gly | Leu | Asn | Ser | Gly | Thr |
|     |     | 165 |     |     |     |     | 170 |     |     |     |     |     | 175 |     |     |
| Gly | Asn | Ile | Gly | Leu | Phe | Asn | Ser | Gly | Thr | Gly | Asn | Val | Gly | Ile | Gly |
|     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |     |
| Asn | Ser | Gly | Thr | Gly | Asn | Trp | Gly | Ile | Gly | Asn | Ser | Gly | Asn | Ser | Tyr |
|     |     | 195 |     |     |     | 200 |     |     |     |     | 205 |     |     |     |     |
| Asn | Thr | Gly | Phe | Gly | Asn | Ser | Gly | Asp | Ala | Asn | Thr | Gly | Phe | Phe | Asn |
|     |     | 210 |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| Ser | Gly | Ile | Ala | Asn | Thr | Gly | Val | Gly | Asn | Ala | Gly | Asn | Tyr | Asn | Thr |
|     |     | 225 |     |     | 230 |     |     |     | 235 |     |     |     | 240 |     |     |

Gly Ser Tyr Asn Pro Gly Asn Ser Asn Thr Gly Gly Phe Asn Met Gly  
 245 250 255  
 Gln Tyr Asn Thr Gly Tyr Leu Asn Ser Gly Asn Tyr Asn Thr Gly Leu  
 260 265 270  
 Ala Asn Ser Gly Asn Val Asn Thr Gly Ala Phe Ile Thr Gly Asn Phe  
 275 280 285  
 Asn Asn Gly Phe Leu Trp Arg Gly Asp His Gln Gly Leu Ile Phe Gly  
 290 295 300  
 Ser Pro Gly Phe Phe Asn Ser Thr Ser Ala Pro Ser Ser Gly Phe Phe  
 305 310 315 320  
 Asn Ser Gly Ala Gly Ser Ala Ser Gly Phe Leu Asn Ser Gly Ala Asn  
 325 330 335  
 Asn Ser Gly Phe Phe Asn Ser Ser Ser Gly Ala Ile Gly Asn Ser Gly  
 340 345 350  
 Leu Ala Asn Ala Gly Val Leu Val Ser Gly Val Ile Asn Ser Gly Asn  
 355 360 365  
 Thr Val Ser Gly Leu Phe Asn Met Ser Leu Val Ala Ile Thr Thr Pro  
 370 375 380  
 Ala Leu Ile Ser Gly Phe Phe Asn Thr Gly Ser Asn Met Ser Gly Phe  
 385 390 395 400  
 Phe Gly Gly Pro Pro Val Phe Asn Leu Gly Leu Ala Asn Arg Gly Val  
 405 410 415  
 Val Asn Ile Leu Gly Asn Ala Asn Ile Gly Asn Tyr Asn Ile Leu Gly  
 420 425 430  
 Ser Gly Asn Val Gly Asp Phe Asn Ile Leu Gly Ser Gly Asn Leu Gly  
 435 440 445  
 Ser Gln Asn Ile Leu Gly Ser Gly Asn Val Gly Ser Phe Asn Ile Gly  
 450 455 460  
 Ser Gly Asn Ile Gly Val Phe Asn Val Gly Ser Gly Ser Leu Gly Asn  
 465 470 475 480  
 Tyr Asn Ile Gly Ser Gly Asn Leu Gly Ile Tyr Asn Ile Gly Phe Gly  
 485 490 495  
 Asn Val Gly Asp Tyr Asn Val Gly Phe Gly Asn Ala Gly Asp Phe Asn  
 500 505 510  
 Gln Gly Phe Ala Asn Thr Gly Asn Asn Asn Ile Gly Phe Ala Asn Thr  
 515 520 525  
 Gly Asn Asn Asn Ile Gly Ile Gly Leu Ser Gly Asp Asn Gln Gly Gly

| 530 |     |     |     | 535 |     |     |     | 540 |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Phe | Asn | Ile | Ala | Ser | Gly | Trp | Asn | Ser | Gly | Thr | Gly |
| 545 |     |     |     | 550 |     |     |     | 555 |     |     | 560 |
| Phe | Asn | Ser | Gly | Thr | Asn | Asn | Val | Gly | Ile | Phe | Asn |
|     |     |     |     | 565 |     |     |     | 570 |     |     | 575 |
| Asn | Val | Gly | Ile | Ala | Asn | Ser | Gly | Thr | Gly | Asn | Trp |
|     |     |     |     | 580 |     |     |     | 585 |     |     | 590 |
| Pro | Gly | Thr | Asp | Asn | Thr | Gly | Ile | Leu | Asn | Ala | Gly |
|     |     |     |     | 595 |     |     |     | 600 |     |     | 605 |
| Gly | Ile | Leu | Asn | Ala | Gly | Asp | Phe | Asn | Thr | Gly | Phe |
|     |     |     |     | 610 |     |     |     | 615 |     |     | 620 |
| Ser | Tyr | Asn | Thr | Gly | Gly | Phe | Asn | Val | Gly | Asn | Thr |
|     |     |     |     | 625 |     |     |     | 630 |     |     | 635 |
| Phe | Asn | Val | Gly | Asp | Thr | Asn | Thr | Gly | Ser | Tyr | Asn |
|     |     |     |     | 640 |     |     |     | 645 |     |     | 650 |
| Asn | Thr | Gly | Phe | Phe | Asn | Pro | Gly | Asn | Val | Asn | Thr |
|     |     |     |     | 655 |     |     |     | 660 |     |     | 665 |
| Thr | Gly | Asp | Phe | Asn | Asn | Gly | Phe | Leu | Val | Ala | Gly |
|     |     |     |     | 670 |     |     |     | 675 |     |     | 680 |
| Gln | Ile | Ala | Ile | Asp | Leu | Ser | Val | Thr | Thr | Pro | Phe |
|     |     |     |     | 685 |     |     |     | 690 |     |     | 695 |
| Glu | Gln | Met | Val | Ile | Asn | Val | Glu | Asn | Val | Met | Thr |
|     |     |     |     | 700 |     |     |     | 705 |     |     | 710 |
| Met | Ile | Thr | Val | Thr | Glu | Ala | Ser | Thr | Val | Glu | Pro |
|     |     |     |     | 715 |     |     |     | 720 |     |     | 725 |
| Leu | Ser | Gly | Leu | Phe | Phe | Phe | Gly | Pro | Val | Asn | Leu |
|     |     |     |     | 730 |     |     |     | 735 |     |     | 740 |
| Leu | Thr | Val | Pro | Thr | Ile | Thr | Leu | Thr | Ile | Gly | Gly |
|     |     |     |     | 745 |     |     |     | 750 |     |     | 755 |
| Val | Pro | Ile | Ser | Ile | Val | Gly | Ala | Leu | Glu | Ser | Arg |
|     |     |     |     | 760 |     |     |     | 765 |     |     | 770 |
| Leu | Lys | Ile | Asp | Pro | Ala | Pro | Gly | Ile | Gly | Asn | Ser |
|     |     |     |     | 775 |     |     |     | 780 |     |     | 785 |
| Ser | Ser | Gly | Phe | Phe | Asn | Ser | Gly | Thr | Gly | Gly | Thr |
|     |     |     |     | 790 |     |     |     | 795 |     |     | 800 |
| Asn | Val | Gly | Gly | Gly | Ser | Ser | Gly | Val | Trp | Asn | Ser |
|     |     |     |     | 805 |     |     |     | 810 |     |     | 815 |



Ala Ile Gly Asn Ser Gly Phe Gln Asn Leu Gly Ser Leu Gln Ser Gly  
 835 840 845  
 Trp Ala Asn Leu Gly Asn Ser Val Ser Gly Phe Phe Asn Thr Ser Thr  
 850 855 860  
 Val Asn Leu Ser Thr Pro Ala Asn Val Ser Gly Leu Asn Asn Ile Gly  
 865 870 875 880  
 Thr Asn Leu Ser Gly Val Phe Arg Gly Pro Thr Gly Thr Ile Phe Asn  
 885 890 895  
 Ala Gly Leu Ala Asn Leu Gly Gln Leu Asn Ile Gly Ser Ala Ser Cys  
 900 905 910  
 Arg Ile Arg His Glu Leu Asp Thr Val Ser Thr Ile Ile Ser Ala Phe  
 915 920 925  
 Cys Gly Ser Ala Ser Asp Glu Ser Asn Pro Gly Ser Val Ser Gln  
 930 935 940

(5) INFORMATION FOR SEQ ID NO:260:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 53 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:260:

GGATGATAT GGGCCATCAT CATCATCATC ACGTGTGGA CATCATCGGG ACC

53

(5) INFORMATION FOR SEQ ID NO:261:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 42 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:261:

CATGATATCA GGGCTGGGTT GGGCGGGGCT CATCTTGAAC GA

42

(5) INFORMATION FOR SEQ ID NO:262:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:202:

GGATCTCTGCA GGTCTGAAAC CACCGAGCG T 31

(2) INFORMATION FOR SEQ ID NO:203:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xii) SEQUENCE DESCRIPTION: SEQ ID NO:203:

CTCTGAAATC AGCGCTGAAA ATCGTCGCG T 31

(2) INFORMATION FOR SEQ ID NO:204:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xiii) SEQUENCE DESCRIPTION: SEQ ID NO:204:

GGATCCAGCG CTGAGCTGAA GACCGATCG T 33

(2) INFORMATION FOR SEQ ID NO:205:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:205:  
GGATATCTGC AGAATTCAGG TTAAAGGCC ATTTCGCA 38

(2) INFORMATION FOR SEQ ID NO:206:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:206:  
CCCATGCGCA CCGACGTGGC CACAAGGGCC 36

(2) INFORMATION FOR SEQ ID NO:207:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:207:  
CTTCATGGRN TTCTCAGGCC GTTAAGGTTC GTGNGG 37

(2) INFORMATION FOR SEQ ID NO:208:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2676 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:208:  
TGGGWAJGG GACGECCTT GTAGCGGCGC ATTAAGCGCG CGGSGTGTGG TGGTACGCG 60

|            |            |            |            |            |             |      |
|------------|------------|------------|------------|------------|-------------|------|
| CAGCGTGACC | GCTACACTTC | CCAGCGCCCT | AGCGCGCGCT | CGTTTCGCTT | TCTTCCTTTC  | 126  |
| CTTTGTCGCC | ACGTTGCGCG | GCTTTCGCCG | TCAGGCTCTA | AATCGGGGCG | TCTCTTTAGG  | 180  |
| GTTCGGATTT | ACGCTTTTTC | GGCAGCTCGA | CCCCAAAANA | CTTGATTAGG | GTATGATTC   | 240  |
| ACGTATGCGG | CGATCGGCTT | GATGACGCTT | TTTTGCGCCT | TTGAGCTTGG | AGTCCAGCTT  | 300  |
| CTTTATATAT | GGACTCTTGT | TCUAAACTGG | AGCAGCCTTC | AGCGCTATCT | CGCTCTATTC  | 360  |
| TTTGTGATTA | TAAGGGATTT | TGCGGATTTT | GCGCTATTGG | TTAAAAAATG | AGCTGATTTT  | 420  |
| ACGAAATTTT | AGGCGGATTT | TTACAGAAAT | ATTAGGCTTT | ACAATTTTCG | GTGCGACTTT  | 480  |
| TGCGCGAAAT | GTGCGGCGAA | CCGCTATTTG | TTTATTTTTT | TAAATGACTT | CAATATGCTA  | 540  |
| TGCGCTCAGG | AATTATTTCT | TAGAAAAACT | CATCGAGCAT | CAATGAAAC  | TGCTATTTAT  | 600  |
| TCATATCAGG | ATTATCAATA | CGATTTTTTT | GAAAGAGCGG | TTTCTCTATT | GAAGGAGAAA  | 660  |
| AGTCAGCGCG | GCAGTTCTAT | AGGATCGGAA | GATCGTGGTA | TGCGTCTCGG | ATTCTGACTC  | 720  |
| GTGCAAGATC | AATACAGCTT | ATTATTTTCC | CGTCTGCGAA | AATAAGGTTA | TGAGTGAGAA  | 780  |
| AATCGACATG | AGTGACACTT | CAATCGGCTG | AGATGCGCAA | AGCTTTATGC | ATTCTTTTCC  | 840  |
| AGACTTTGTC | AACAGCGCAG | CGCTTAAGCT | CGCTATCAAA | ATCACTCGCA | TCAGCCAAAC  | 900  |
| CGTTATTCAT | TCTGATTTTC | GCTGAGCGGA | GACGAATTAC | GGGATCGCTG | TTAAAGGAGC  | 960  |
| AATTACAAAG | AGCAATCGAA | TGCAACCGGC | CGAGGAGGAC | TGCGAGGCGA | TCAGCATATAT | 1020 |
| TTTCACTCTA | ATGAGGATAT | TCTCTTAATA | CGTGAGATGC | TGTTTTGCGG | GGGATGCGAG  | 1080 |
| TGGGAGATTA | CGATGAGTCA | TGAGGAGTAC | GGATAAAATG | CTTGATGCTC | GGAGAGGCGA  | 1140 |
| TAAATTCGCT | GAGGCGCTTT | AGCTGACCTT | TCTCATCTGT | AGCATCATTC | GCAGCGCTAC  | 1200 |
| CTTTGCGCAT | TTTGAGGAAC | AATCTGCGCG | CATCGGCGCT | CGCATACAT  | CGATATATTC  | 1260 |
| TGCGAGCTGA | TTGCGGAGAA | TTATCGGAGC | CGCATTTATA | CGCATATAAA | TCAGCATCTCA | 1320 |
| TGTTGGAATT | TATGCGCGCG | CTAGAGCGAG | AGCTTTGCGG | TTGAAGTTGG | CTCTATACAC  | 1380 |
| CGCTGCTATT | AGCTTTATTC | TAGGAGGACT | GTTTTATTTG | TCTGAGCGAA | AATCGCTTAA  | 1440 |
| CGTGAATTTT | CGTTGAGCTG | AGCGTCAGAC | CGCTGAGAAA | AGATCGAGAG | ATCTTTCTGA  | 1500 |
| GATGCTTTTT | TTTGGCGGCT | AATCTGCTTC | TTTCAAGCAA | AAAAGGAGCG | CGTTCGAGCG  | 1560 |
| GCGCTTTTCT | TGAGGAGTCA | AGGCTGACCA | AGCTTTTCTC | CGAGGATAGC | TGCTTTGAGC  | 1620 |
| AGAGCGCAGA | TAGCAATATC | TGCTCTCTCA | GTGAGGCTAT | AGTTAGGCGA | CGCTTTCAAG  | 1680 |

|             |            |            |            |            |             |      |
|-------------|------------|------------|------------|------------|-------------|------|
| AACCTCTGTAG | CACCGGCTAC | ATACCTGAGT | CTGCTAATTC | TGTTACCACT | GGCTGTGCGC  | 1740 |
| AGTGGCGATA  | AGTGTGTCTT | TACCGGGTTG | GACTCAAGAC | GATATTTACG | GGATTAAGGCG | 1800 |
| CAGGCGTCGG  | GCTGAACGGG | GGCTTCGTGC | ACACAGCCCA | GCTTGGAGCG | AACGACCTAC  | 1860 |
| ACCGAAGTGA  | GATACCTACA | GCCTGAGCTA | TGAGAAAGCG | CCACGCTTCC | CGAAGCGAGA  | 1920 |
| AAGCGGGAAC  | GGTATCCGGT | AAGCGGCGAG | GTCGGAACAG | GAGAACGCGC | GAGGAGGCTT  | 1980 |
| CCAGGGGGA   | ACGCGTGGTA | TGTTTATAGT | CGTGTGCGGT | TTGACCGAGT | CTGAGTTGAG  | 2040 |
| CTCTGATTTT  | TGTGATGCTC | GTGAGGCGGG | GAGAGCTTAT | GGAAAAACGC | GAGCAAGCGG  | 2100 |
| GGCTTTTTCG  | GGTTCCTGCG | CTTTTGCTGG | CGTTTTGCTC | ACATGTTCTT | TCTGACGTTA  | 2160 |
| TGCGCTGATT  | CTGTGGATAA | CGGTATTACG | GGCTTGGAGT | GAGCTGATAC | CGCTCGCCGC  | 2220 |
| AGCGGAAGGA  | CTGAGCGGAG | CGACTCAGTC | AGCGAGGAGG | CGGAAGAGCG | GCTGATGCGG  | 2280 |
| TATTTTCTCC  | TTAGGCGATC | GTGCGGTATT | TCAGACGGCA | TATATGTCGC | ACTCTCACTA  | 2340 |
| CAATCTGCTC  | TCATGCGGCA | TAGTTAAGGC | AGTATGACCT | CGGCTATCGC | TACCTGACTG  | 2400 |
| GGTGTGCGCT  | GCGCGCGGAC | ACCGCGGAAC | ACCGGCTGAC | GGGCTCTGAG | GGGCTGTGCT  | 2460 |
| GCTCTCGGCA  | TGGCTTACCA | GACAGGCTGG | GACGCTCTCC | GGGAGCTGCA | TGTTGTGAGG  | 2520 |
| GTTTTCACCG  | TCATCAGGGA | AGCGCGGAGG | GGAGCTGCGG | TAAAGCTCAT | CAGCGTGCTC  | 2580 |
| GTAAGGCGAT  | TCAGAGATGT | GTGCTGTGTC | ATCGGCTGTC | AGCTGCTTGA | GTTTCTCCAG  | 2640 |
| AGCGTTAAT   | GTGCGGCTTC | TGATTAAGCG | GGCATGTTTA | AGCGCGGTTT | TTTCTGTGTT  | 2700 |
| GGTCACTGAT  | GGCTCGGTGT | AAGCGGGATT | TCTGTTCATC | GGGCTAATGA | TACGATGAAA  | 2760 |
| ACGAGAGAGG  | ATGCTCTGGA | TACGCTTAC  | TGATGATGAA | CATGCGGGGT | TACTCGAAGC  | 2820 |
| TTGTGAGGCT  | AAACACTGCG | CGGATGCGAT | GGCGCGGAC  | CAGAGAGAAA | TCAGTCAGGG  | 2880 |
| TGATGCGCG   | CGCTTCGTTA | ATGAGATGCT | AGGTGTTGCA | CAGGTTATCC | AGCAGATCC   | 2940 |
| TGGATCCAG   | ATCGGAGGCA | TATGCTGCA  | GGCGCTGAC  | TTCCGCTGTT | CGAGACTTTA  | 3000 |
| CGAAACACCG  | AGACCGAAGA | CGATCATGCT | TGTTGCTCGG | GTGCGAGGCG | TTTTCAGGCA  | 3060 |
| CGACTGCGCT  | CAGTTGCGCT | CGGCTATGCG | TGATTCATTC | TGCTAAGCAG | TAGGGTAAGC  | 3120 |
| CGCGGACCTT  | AGCGGGGTCC | TGACCTACAG | GACGACGATC | ATGCGGACCT | GTGCGGCGCG  | 3180 |
| CATGCGGCGG  | ATATGCGGCT | GCTTTCGCGC | GAAACGTTTC | GTGCGGAGAC | CAGTCAGGAA  | 3240 |
| GGCTGAGGCG  | AGCGGCTGCA | AGATTCTGAA | TACGCGAGGC | GACGAGCGGA | TCATCTCTCG  | 3300 |
| GGTTCAGCGA  | AGCGGCTGCT | CGCGGAAAT  | GACGCGAGGC | GCTGCGCGCA | GCTGTCTTAC  | 3360 |

|            |            |           |            |            |            |      |
|------------|------------|-----------|------------|------------|------------|------|
| GAGTGGCATG | ATAAAGAGA  | CAGTCATAG | TGCGGGAGCG | ATAGTCATCG | CGGCGGCCCA | 3420 |
| CCGGAGGAG  | CTGACTGGGT | TGAGGCTCT | CAAGGGCATC | GATCGAGTC  | CGGATGCTTA | 3480 |
| ATGAGTGAGC | TAACTACAT  | TAAATGCGT | GCGGTCACTG | CCCGCTTTC  | AGTCGGGAAA | 3540 |
| GCTGTGATGC | CAGTCGATT  | ATGAATCGG | CGAACGCGG  | GCGACGGGG  | GTTTGGGTAT | 3600 |
| TGGGGGAG   | GCTGTTTTT  | CTTTACACA | CTGAGCGGG  | GAACAGTTGA | TGGGCTTCA  | 3660 |
| CGGCTGGG   | CTGAGAGAT  | TGCGGAGAG | GCTGAGGCT  | GCTTGGGCG  | AGCAGGGGAA | 3720 |
| ATGCTGTTT  | GATGCTGAT  | AACGGCGGA | TATAACATTA | GCTGCTTGG  | GATCTGCTG  | 3780 |
| ATCCGACTAC | CGAGATATC  | GCAGGAGGC | GCAGCGGGA  | CTGGTAATG  | GCGGCGATTG | 3840 |
| CGCGCGCG   | CATCTGATG  | TGCGAGCGA | GCATCGCGT  | AGGAGCGAG  | CGCTGATTA  | 3900 |
| GCATTTCAT  | GCTTGTGTA  | AAAGCGACA | TGGGACTCG  | GTCGCTTGG  | CTTTCGCTA  | 3960 |
| TGCGCTGAT  | TGATGCGG   | GAGAGATAT | TATGCGGCG  | AGGAGAGCG  | AGAGCGGCG  | 4020 |
| AGACGAGCT  | TATGCGCG   | GCTAACAGG | CGATTGCTG  | CGGCGCAAT  | CGGAGCAGAT | 4080 |
| GTTCGAGCG  | CAGTCGATA  | CGGCTTCAT | GGAGAGAAAT | ATGCTGCTG  | ATGCTGCTG  | 4140 |
| GCTGAGAGC  | ATCAGAGAT  | AACGCGGAA | CTTGTGCTA  | CGCGGCTTC  | AGAGCAATCG | 4200 |
| CATCTGCTG  | ATCAGAGAG  | TAGTTAATG | TGAGCGGCT  | AGCGGCTTC  | CGGAGAGAT  | 4260 |
| TGTCGAGCG  | CGCTTTCAG  | GCTTCGAGC | CGCTTCTTC  | TAGCATGAG  | AGGAGAGCG  | 4320 |
| TGCGAGGAG  | TGATCGGCG  | CGAGATTAA | TGCGGCGAG  | AATTGCGAG  | GCGGCTGCA  | 4380 |
| GCGCGAGCT  | CGAGCTGCA  | AGCGCAATC | CGAGGAGCT  | TTCGCGCG   | AGTGTGCTG  | 4440 |
| CGAGGCGCT  | CGAGGATA   | TTCAGCTCG | CGAGGCGCG  | TTCAGCTTC  | TGCGGCTTC  | 4500 |
| TGCGAGAGC  | GTCGCTGCG  | TGCTTCAGC | CGCGGAGAG  | GCTTCGATA  | AGAGAGCGG  | 4560 |
| CATCTCTCG  | GAGTCGAT   | AGCTTACG  | GTTTCAGAT  | CGAGCGGCG  | AGTGTGCTG  | 4620 |
| CTTCGCGCG  | CTATCGAG   | ATACCGGAG | AGCTTTCGCG | CGAGGAGCG  | AGTGTGCGG  | 4680 |
| TCTGAGGCT  | CTGCTTATG  | CGAGTCGCG | ATAGGAGAG  | AGCGGAGAG  | TGCTTTCAG  | 4740 |
| CGCTTTCAG  | CGCGCGCG   | AGGAGGCT  | CGAGGAGCG  | AGAGGAGCG  | CGAGGCTCG  | 4800 |
| CGCGGAGCG  | CGCTTTCAG  | CGAGGAGCG | CGAGGAGCG  | CGAGGAGCG  | CGAGGAGCG  | 4860 |
| CGAGGAGCG  | CGCTTTCAG  | CGAGGAGCG | CGAGGAGCG  | CGAGGAGCG  | CGAGGAGCG  | 4920 |
| CGCGGAGCG  | CGCTTTCAG  | CGAGGAGCG | CGAGGAGCG  | CGAGGAGCG  | CGAGGAGCG  | 4980 |

|            |            |            |            |            |            |      |
|------------|------------|------------|------------|------------|------------|------|
| AATTATACG  | ACTCAGTATA | GGGGATTGT  | GAGCGATAA  | CAATTCCCT  | CTAGAAATAA | 5040 |
| TTTTGTTTAA | TTTAAAGAG  | GAGATATACA | TATGCGCCAT | CATCATCATC | ATCAGCTGAT | 5100 |
| CGACATATC  | GCGACGAGC  | CCACATCTC  | GGAACAGGC  | GCGCGGAGG  | CGGTCCAGC  | 5160 |
| GGCGCGGAT  | AGCTCGATG  | AGTCCGCGT  | CGCTCGGTC  | ATTGAGGAG  | ACATCGCGCT | 5220 |
| GGACAGGCG  | GCGAAGATC  | CTACCGCAT  | CAAGCTGAA  | GTCTCGTTA  | AGATGAGGCT | 5280 |
| GCGCAACCG  | AGCGGCTGA  | AACACCGAG  | CGTTCGCT   | GAAACGGCG  | CGGCGGCGG  | 5340 |
| TACTCTGCG  | ACTACCGCG  | CGTCTCGG   | GCTGAGCTG  | GCGAGAGCG  | GTACCGGCT  | 5400 |
| GCTCTACCG  | CTGTTCACT  | TGTCGGTCC  | GCTCTTTC   | GAGCGGATC  | CGACCTCAC  | 5460 |
| GATCAGCTC  | CAGGCGAGC  | GTCTCGGTC  | CGGATCGAG  | CAGCGGCGG  | CGGCGGCGT  | 5520 |
| CAACATTGG  | CGCTCGAGC  | CTTCTCTCT  | GGAAGGCTG  | ATGCGCGCG  | ACAGCGGCT  | 5580 |
| GATGAAGATC | CGCTAGGCA  | TCTCGCTCA  | GCGGCTAAC  | TACAGCTCG  | CGGATCTAG  | 5640 |
| CGAGCAGCT  | AAGCTGAAG  | GAAATCTCT  | CGCGCGGAG  | TACCGGCG   | CGATCMAAC  | 5700 |
| CTGAGCAGC  | CGGAGAGCG  | CTCGGCTTA  | CGCGCGGTC  | AACCTCGCG  | CGGCGGCGT  | 5760 |
| AGTTCGCTG  | CAGGCTCGG  | AGGCTCTCG  | TGACATCTT  | TGCTTACCT  | AGTACTCTG  | 5820 |
| CAAGCAAGAT | CGCGAGGCT  | GCGGAGGTC  | CGCGGCTTC  | GCGACCGCG  | TGCACTTCTG | 5880 |
| GCGGCTCGG  | CTGCGGCTG  | GTGAGAGCG  | CAAGCGGCG  | ATGTCGAGC  | CTTTCGCGGA | 5940 |
| GACAGCGCG  | TGCTGCTCT  | ATATCGCAT  | CACTTCTCT  | GACAGCGCA  | GTCAACGAG  | 6000 |
| ACTTCGCGG  | CGCGAGTAA  | GCAATAGCT  | TGCGAATTC  | TGCTTGGCG  | ACCGCGAAG  | 6060 |
| CATTACGCG  | GCGCGGCTG  | GCTTCGATC  | GAAACCGCG  | GCGACCGAG  | CGATTTGAT  | 6120 |
| GATGACGCG  | CGCGGCTCG  | AGGCTACCT  | GATCATCAC  | TAGAGTAGG  | CGATCTCAA  | 6180 |
| CAACCGGGA  | AGGAGCGCG  | CGACCGGCG  | GAGTTTGA   | CGATTTCTG  | ACTGCGGAT  | 6240 |
| CACCGAGCG  | AACAGGCTT  | CGTTCTCGA  | CGAGTTCT   | TTCAGCGCG  | TGCTCGCGG  | 6300 |
| GTGCTGAG   | TTCTCTGAG  | CGTTCTCTG  | GAGATTCT   | AGGCTGAGA  | TGAGACCGA  | 6360 |
| TGCGCTTAC  | CTCGCGGCG  | AGGAGGTA   | TTTGGAGCG  | ATCTCGGCG  | AGCTGAAAG  | 6420 |
| CGAGATGAG  | GAGCTGAT   | GAGCGGAG   | TTCCTGAG   | GAGCTGAG   | CGGCGGCGG  | 6480 |
| GCGAGCGCG  | GCGAGGCTG  | CGGCTCTCG  | CTTCTAGAG  | GCGCGGATA  | AGCAGAGCA  | 6540 |
| GGAAGCTAG  | GAGATCTGA  | CGAGATCTG  | TGAGGCGCG  | GTCTAATCT  | CGAGGCGCG  | 6600 |
| CGAGGAGCG  | CAGCAAGCG  | TGCTCTGCG  | ATGCGCTTT  | GTGCGGAGA  | CGGCGGCTC  | 6660 |

```

GGGGGGCTGG ACCGCTGCAG GGGGATCGGG ACCGCGACAG CTCTGTGAGG CCGGACCAAG 6720
GGCGGCGGGG AACACGGCTGA ATGCGGAGGC GGCGGATCCG AACGGAGGAC CTGCGGCGGG 6760
GGACCGGAGC GCGCGCGCGC CACCTGTGAT TGCCCGAAGC GCGCGGCAAC CTGTGCGGAT 6840
GGACACCGCG GTTGGAGGAT TCAGCTTCGG GCTGCTTGGT GGTGTGATGG AGTCTGAGGC 6900
CGCGCACTTC GACTACGCTT CAGGACTGCT CAGCAAAAGC ACCGCGGAGC GCGCATTTCC 6960
GGGACAGCGG CGCGCGGCTG CCAATGACAC CCGTATGCTG CTGCGCGCGC TAGACCAAAA 7020
GTTTTAGCGC AGCGCGGAGC CCACCGACTG CAAGCGCGCG GCGGCTTGGG GCTGCGACAT 7080
GGGTGACTTC TGATGCGGCT ACCGCGGACG CGGATCAAC CAGGAGACCG TCTGCGTTGA 7140
CGCCAGCGGG GTGCTCGGAA GCGCGTGCTA TTAGGAAGTC AAGTTCACCG ATCGGAGTAA 7200
GGCGACGGCG CAGATCTGGA CGGCGCTAAT GGGCTCGCGC GCGCGGAGCG CAGCGGACCG 7260
CGGCGCGGCT CAGCGCTGCT TTGTGCTATG GCTCGGAGCC GCGACAGACG CGGTGAGACG 7320
GGGCGCGCGC AAGCGCGTGG CCGAATGCGT CCGCGCTTTC GTGCGCGCGC GCGCGCGCGC 7380
GGCACCGGCT GTCGAGAGC CCGCTCGGCG GCGCGCGCGC GCGCGCGGAG TCGCTCTTAC 7440
CGCGACGACA CGGACGCGCG AGCGGACCTT ACCGCGCTGA GATTCTGCA GATATCCATC 7500
ACATCGCGCG CGGCTCGAGC ACCGCGCGCA CCAGGACTGA GATCGCGCTG CTAACAAAGC 7560
CGCAAGGAGG GCTGAGTTGG CTGCTGCGCG CGGTGAGGAA TAAGTAGCAT AACCGCTTGG 7620
GCGCTCTAAA CGGCTCTTGA GCGGTCTTTT GCTGAAGGGA GCACTATAT CCGGAT 7676

```

## (2) INFORMATION FOR SEQ ID NO:209:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 802 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) SEQUENCE DESCRIPTION: SEQ ID NO:209:

```

Met Gly His His His His His Val Ile Asp Ile Ile Gly Thr Ser
1           5           10
Pro Thr Ser Thr Glu Glu Ala Ala Ala Glu Ala Val Glu Arg Ala Arg
20          25          30

```



Asp Ser Val Asp Asp Ile Arg Val Ala Arg Val Ile Gln Gln Asp Met  
 35 40 45  
 Ala Val Asp Ser Ala Gly Lys Ile Thr Tyr Arg Ile Lys Leu Gln Val  
 50 55 60  
 Ser Phe Lys Met Arg Pro Ala Gln Pro Arg Gly Ser Lys Pro Pro Ser  
 65 70 75 80  
 Gly Ser Pro Gln Thr Gly Ala Gly Ala Gly Thr Val Ala Thr Thr Pro  
 85 90 95  
 Ala Ser Ser Pro Val Thr Leu Ala Gln Thr Gly Ser Thr Leu Leu Tyr  
 100 105 110  
 Pro Leu Phe Asn Leu Trp Gly Pro Ala Phe His Gln Arg Tyr Pro Asn  
 115 120 125  
 Val Thr Ile Thr Ala Gln Gly Thr Gly Ser Gly Ala Gly Ile Ala Gln  
 130 135 140  
 Ala Ala Ala Gly Thr Val Asn Ile Gly Ala Ser Asp Ala Tyr Leu Ser  
 145 150 155 160  
 Gln Gly Asp Met Ala Ala His Lys Gly Leu Met Asn Ile Ala Leu Ala  
 165 170 175  
 Ile Ser Ala Gln Gln Val Asn Tyr Asn Leu Pro Gly Val Ser Gln His  
 180 185 190  
 Leu Lys Leu Asn Gly Lys Val Leu Ala Ala Met Tyr Gln Gly Thr Ile  
 195 200 205  
 Lys Thr Trp Asp Asp Pro Gln Ile Ala Ala Leu Asn Pro Gly Val Asn  
 210 215 220  
 Leu Pro Gly Thr Ala Val Val Pro Asn His Arg Ser Asp Gly Ser Gly  
 225 230 235 240  
 Asp Thr Phe Leu Phe Thr Gln Tyr Leu Ser Lys Gln Asp Pro Gln Gly  
 245 250 255  
 Trp Gly Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val  
 260 265 270  
 Pro Gly Ala Leu Gly Gln Asn Gly Asn Gly Gly Met Val Thr Gly Cys  
 275 280 285  
 Ala Gln Thr Pro Gly Cys Val Ala Tyr Ile Gly Ile Ser Phe Leu Asp  
 290 295 300  
 Gln Ala Ser Gln Arg Gly Asn Gly Gln Ala Gln Leu Gly Asn Ser Ser  
 305 310 315 320  
 Gly Asn Phe Leu Leu Pro Asp His Gln Ser Thr Gln Ala Ala Ala



Gly Asp Pro Pro Phe Pro Gly Glu Pro Pro Pro Val Ala Asn Asp Thr  
 625 630 635 640  
 Arg Ile Val Leu Gly Arg Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu  
 645 650 655  
 Ala Thr Asp Ser Lys Ala Ala Ala Arg Leu Gly Ser Asp Met Gly Glu  
 660 665 670  
 Phe Tyr Met Pro Tyr Pro Gly Thr Arg Ile Asn Glu Glu Thr Val Ser  
 675 680 685  
 Leu Asp Ala Asn Gly Val Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys  
 690 695 700  
 Phe Ser Asp Pro Ser Lys Pro Asn Gly Gln Ile Thr Thr Gly Val Ile  
 705 710 715 720  
 Gly Ser Pro Ala Ala Asn Ala Pro Asp Ala Gly Pro Pro Gln Arg Thr  
 725 730 735  
 Phe Val Val Trp Leu Gly Thr Ala Asn Asn Pro Val Asp Lys Gly Ala  
 740 745 750  
 Ala Lys Ala Leu Ala Glu Ser Ile Arg Pro Leu Val Ala Pro Pro Pro  
 755 760 765  
 Ala Pro Ala Pro Ala Pro Ala Glu Pro Ala Pro Ala Pro Ala Pro Ala  
 770 775 780  
 Gly Glu Val Ala Pro Thr Pro Thr Thr Pro Thr Pro Gln Arg Thr Leu  
 785 790 795 800  
 Pro Ala

## CLAIMS

We claim:

1. A polypeptide comprising an antigenic portion of a soluble *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen has an N-terminal sequence selected from the group consisting of:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu (SEQ ID NO: 115);
- (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser (SEQ ID NO: 116);
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg (SEQ ID NO: 17);
- (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro (SEQ ID NO: 118);
- (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val (SEQ ID NO: 119);
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro (SEQ ID NO: 120);
- (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Thr-Ala-Ala-Ser-Pro-Pro-Ser (SEQ ID NO: 121);
- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly (SEQ ID NO: 122);
- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Leu-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn (SEQ ID NO: 123); and
- (j) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID NO: 131)

wherein Xaa may be any amino acid.

2. A polypeptide comprising an immunogenic portion of an *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen has an N-terminal sequence selected from the group consisting of:

- (a) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID NO: 124) and
- (b) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID NO: 132), wherein Xaa may be any amino acid.

3. A polypeptide comprising an antigenic portion of a soluble *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID NOS: 1, 2, 4-10, 13-25, 52, 94 and 96, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 1, 2, 4-10, 13-25, 52, 94 and 96 or a complement thereof under moderately stringent conditions.

4. A polypeptide comprising an antigenic portion of a *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID NOS: 26-51, 133, 134, 158-178 and 196, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 26-51, 133, 134, 158-178 and 196 or a complement thereof under moderately stringent conditions.

5. A DNA molecule comprising a nucleotide sequence encoding a polypeptide according to any one of claims 1-4.

6. A recombinant expression vector comprising a DNA molecule according to claim 5.

7. A host cell transformed with an expression vector according to claim 6.

8. The host cell of claim 7 wherein the host cell is selected from the group consisting of *E. coli*, yeast and mammalian cells.

9. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting a biological sample with one or more polypeptides according to any of claims 1-4; and

(b) detecting in the sample the presence of antibodies that bind to at least one of the polypeptides, thereby detecting *M. tuberculosis* infection in the biological sample.

10. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting a biological sample with a polypeptide having an N-terminal sequence selected from the group consisting of sequences provided in SEQ ID NO: 129 and 130; and

(b) detecting in the sample the presence of antibodies that bind to at least one of the polypeptides, thereby detecting *M. tuberculosis* infection in the biological sample.

11. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting a biological sample with one or more polypeptides encoded by a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198; and

(b) detecting in the sample the presence of antibodies that bind to at least one of the polypeptides, thereby detecting *M. tuberculosis* infection in the biological sample.

12. The method of any one of claims 9-11 wherein step (a) additionally comprises contacting the biological sample with a 38 kD *M. tuberculosis* antigen and step (b) additionally comprises detecting in the sample the presence of antibodies that bind to the 38 kD *M. tuberculosis* antigen.

13. The method of any one of claims 9-11 wherein the polypeptide(s) are bound to a solid support.

14. The method of claim 13 wherein the solid support comprises nitrocellulose, latex or a plastic material.

15. The method of any one of claims 9-11 wherein the biological sample is selected from the group consisting of whole blood, serum, plasma, saliva, cerebrospinal fluid and urine.

16. The method of claim 15 wherein the biological sample is whole blood or serum.

17. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the sample with at least two oligonucleotide primers in a polymerase chain reaction, wherein at least one of the oligonucleotide primers is specific for a DNA molecule according to claim 5; and

(b) detecting in the sample a DNA sequence that amplifies in the presence of the oligonucleotide primers, thereby detecting *M. tuberculosis* infection.

18. The method of claim 17, wherein at least one of the oligonucleotide primers comprises at least about 10 contiguous nucleotides of a DNA molecule according to claim 5.

19. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the sample with at least two oligonucleotide primers in a polymerase chain reaction, wherein at least one of the oligonucleotide primers is specific for a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198; and

(b) detecting in the sample a DNA sequence that amplifies in the presence of the first and second oligonucleotide primers, thereby detecting *M. tuberculosis* infection.

20. The method of claim 19, wherein at least one of the oligonucleotide primers comprises at least about 10 contiguous nucleotides of a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198.

21. The method of claims 17 or 19 wherein the biological sample is selected from the group consisting of whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid and urine.

22. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the sample with one or more oligonucleotide probes specific for a DNA molecule according to claim 5; and

(b) detecting in the sample a DNA sequence that hybridizes to the oligonucleotide probe, thereby detecting *M. tuberculosis* infection.



23. The method of claim 22 wherein the probe comprises at least about 15 contiguous nucleotides of a DNA molecule according to claim 5.

24. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the sample with one or more oligonucleotide probes specific for a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198; and

(b) detecting in the sample a DNA sequence that hybridizes to the oligonucleotide probe, thereby detecting *M. tuberculosis* infection.

25. The method of claim 24 wherein the oligonucleotide probe comprises at least about 15 contiguous nucleotides of a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198.

26. The method of claims 22 or 24 wherein the biological sample is selected from the group consisting of whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid and urine.

27. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the biological sample with a binding agent which is capable of binding to a polypeptide according to any one of claims 1-4; and

(b) detecting in the sample a protein or polypeptide that binds to the binding agent, thereby detecting *M. tuberculosis* infection in the biological sample.

28. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

- (a) contacting the biological sample with a binding agent which is capable of binding to a polypeptide having an N-terminal sequence selected from the group consisting of sequences provided in SEQ ID NO: 129 and 130; and
- (b) detecting in the sample a protein or polypeptide that binds to the binding agent, thereby detecting *M. tuberculosis* infection in the biological sample.

29. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

- (a) contacting the biological sample with a binding agent which is capable of binding to a polypeptide encoded by a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198; and
- (b) detecting in the sample a protein or polypeptide that binds to the binding agent, thereby detecting *M. tuberculosis* infection in the biological sample.

30. The method of any one of claims 27-29 wherein the binding agent is a monoclonal antibody.

31. The method of any one of claims 27-29 wherein the binding agent is a polyclonal antibody.

32. A diagnostic kit comprising:

- (a) one or more polypeptides according to any of claims 1-4; and
- (b) a detection reagent.

33. A diagnostic kit comprising:

- (a) one or more polypeptides having an N-terminal sequence selected from the group consisting of sequences provided in SEQ ID NO: 129 and 130; and
- (b) a detection reagent.

34. A diagnostic kit comprising:
- (a) one or more polypeptides encoded by a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198; and
- (b) a detection reagent.
35. The kit of any one of claims 32-34 wherein the polypeptide(s) are immobilized on a solid support.
36. The kit of claim 35 wherein the solid support comprises nitrocellulose, latex or a plastic material.
37. The kit of any one of claims 32-34 wherein the detection reagent comprises a reporter group conjugated to a binding agent.
38. The kit of claim 37 wherein the binding agent is selected from the group consisting of anti-immunoglobulins, Protein G, Protein A and lectins.
39. The kit of claim 37 wherein the reporter group is selected from the group consisting of radioisotopes, fluorescent groups, luminescent groups, enzymes, biotin and dye particles.
40. A diagnostic kit comprising at least two oligonucleotide primers, at least one of the oligonucleotide primers being specific for a DNA molecule according to claim 5,

41. A diagnostic kit according to claim 40, wherein at least one of the oligonucleotide primers comprises at least about 10 contiguous nucleotide of a DNA molecule according to claim 5.

42. A diagnostic kit comprising a at least two oligonucleotide primers, at least one of the primers being specific for a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198.

43. A diagnostic kit according to claim 42, wherein at least one of the oligonucleotide primers comprises at least about 10 contiguous nucleotide of a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198.

44. A diagnostic kit comprising at least one oligonucleotide probe, the oligonucleotide probe being specific for a DNA molecule according to claim 5.

45. A kit according to claim 44, wherein the oligonucleotide probe comprises at least about 15 contiguous nucleotides of a DNA molecule according to claim 5.

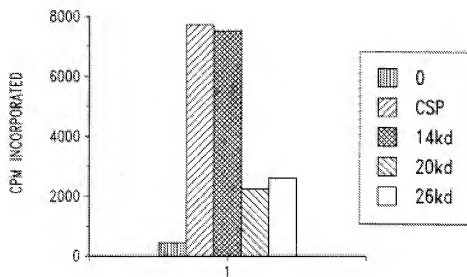
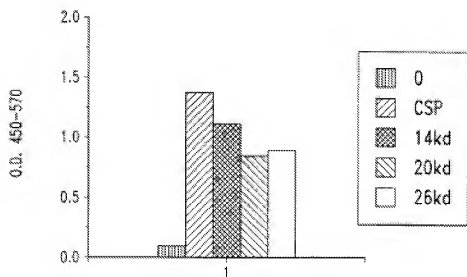
46. A diagnostic kit comprising at least one oligonucleotide probe, the oligonucleotide probe being specific for a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198.

47. A kit according to claim 46, wherein the oligonucleotide probe comprises at least about 15 contiguous nucleotides of a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198.

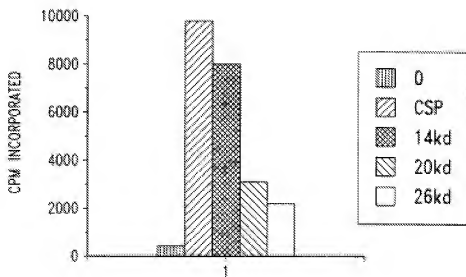
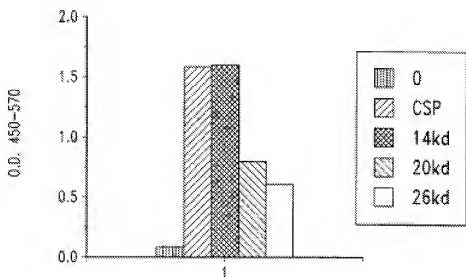
48. A monoclonal antibody that binds to a polypeptide according to any of claims 1-4.

49. A polyclonal antibody that binds to a polypeptide according to any of claims 1-4.
50. A fusion protein comprising two or more polypeptides according to any one of claims 1-4.
51. A fusion protein comprising one or more polypeptides according to any one of claims 1-4 and ESAT-6 (SEQ ID NO: 99).
52. A fusion protein comprising a polypeptide having an N-terminal sequence selected from the group of sequences provided in SEQ ID NOS: 129 and 130.
53. A fusion protein comprising one or more polypeptides according to any one of claims 1-4 and the *M. tuberculosis* antigen 38 kD (SEQ ID NO: 150).
54. A diagnostic kit comprising:
- (a) one or more fusion proteins according to any one of claims 50-53; and
  - (b) a detection reagent.

1/13

*Fig. 1A-1**Fig. 1A-2*

2/13

*Fig. 1B-1**Fig. 1B-2*

3/13

M 1 2 3 4 5

68-  
43-  
29-  
18-  
14-

Fig. 2B

M 1 2 3 4 5

97-  
68-  
43-  
29-  
18-  
14-

Fig. 2D

M 1 2 3 4 5

97-  
68-  
43-  
29-  
18-  
14-

Fig. 2A

M 1 2 3 4 5

97-  
68-  
43-  
29-  
18-  
14-

Fig. 2C



4/13

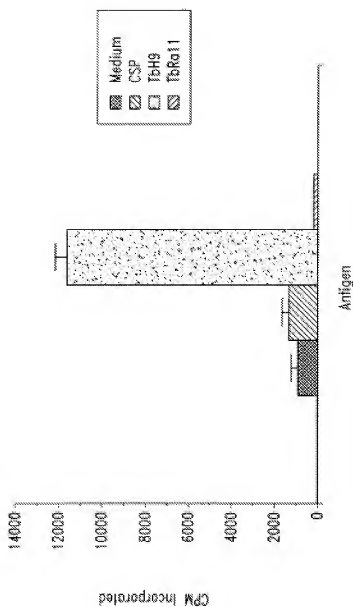
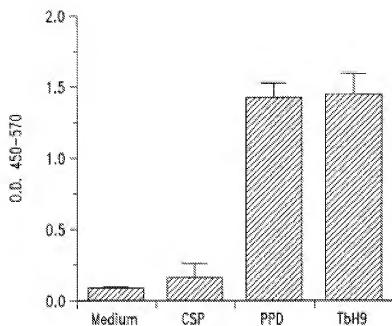


Fig. 3A

5/13

*Fig. 3B*

6/13

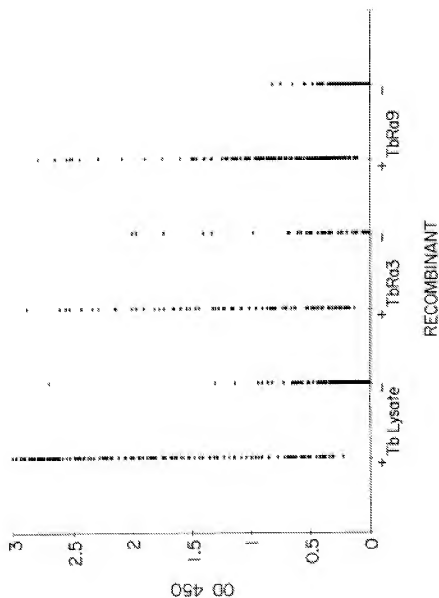


Fig. 4

7/13

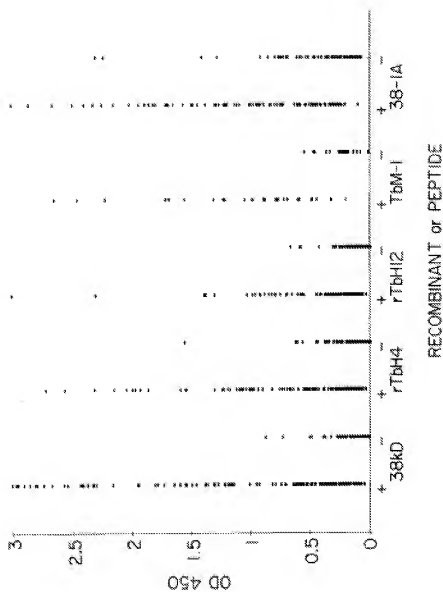


Fig. 5

8/13

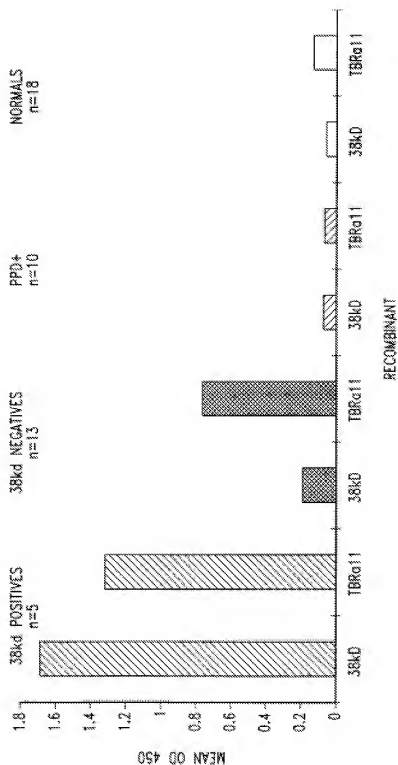


Fig. 6

9/13

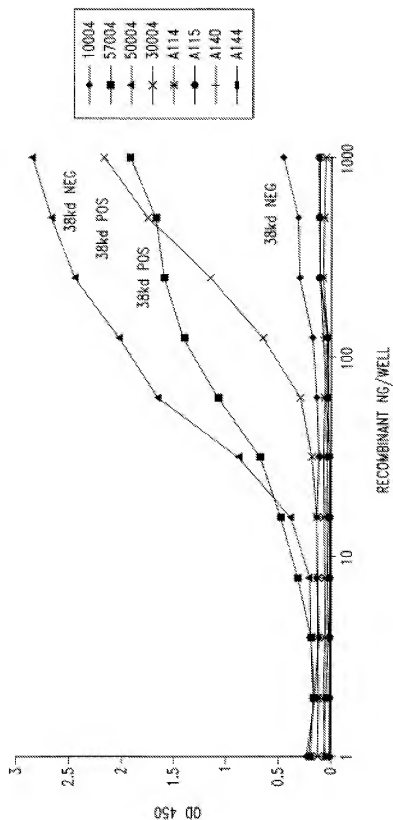


Fig. 7

10/13

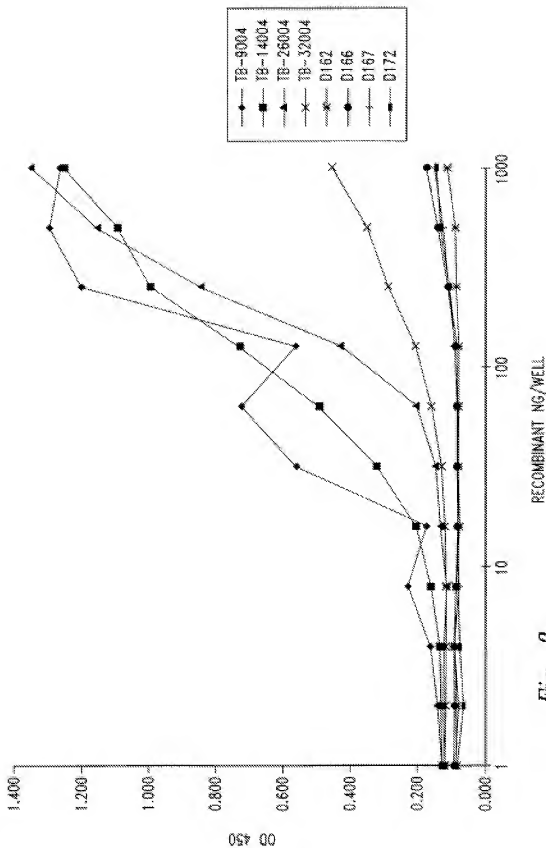


Fig. 8

11/13

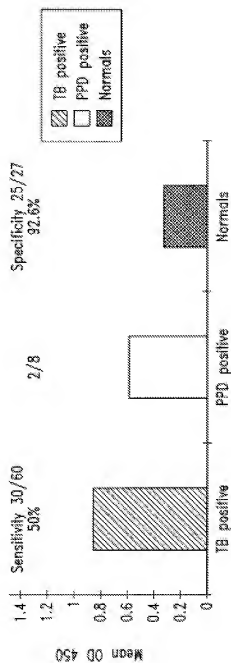


Fig. 9



12/13

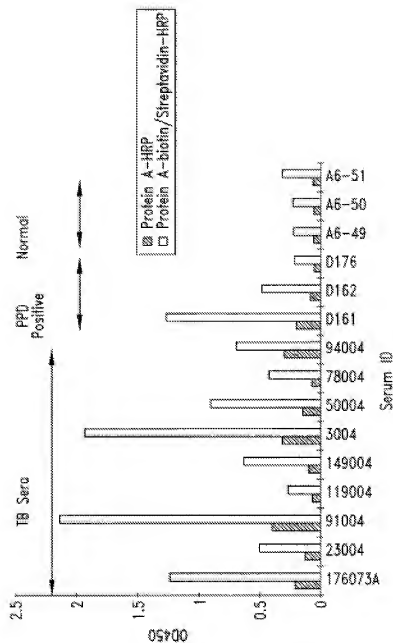


Fig. 10

13/13

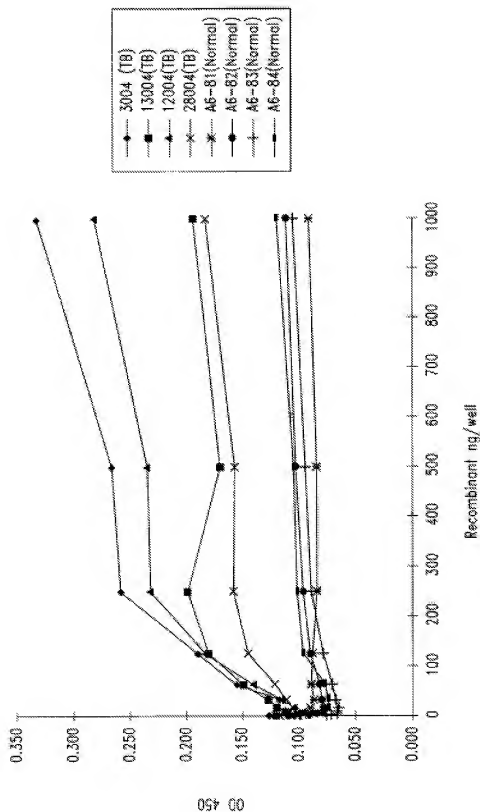


Fig. 11